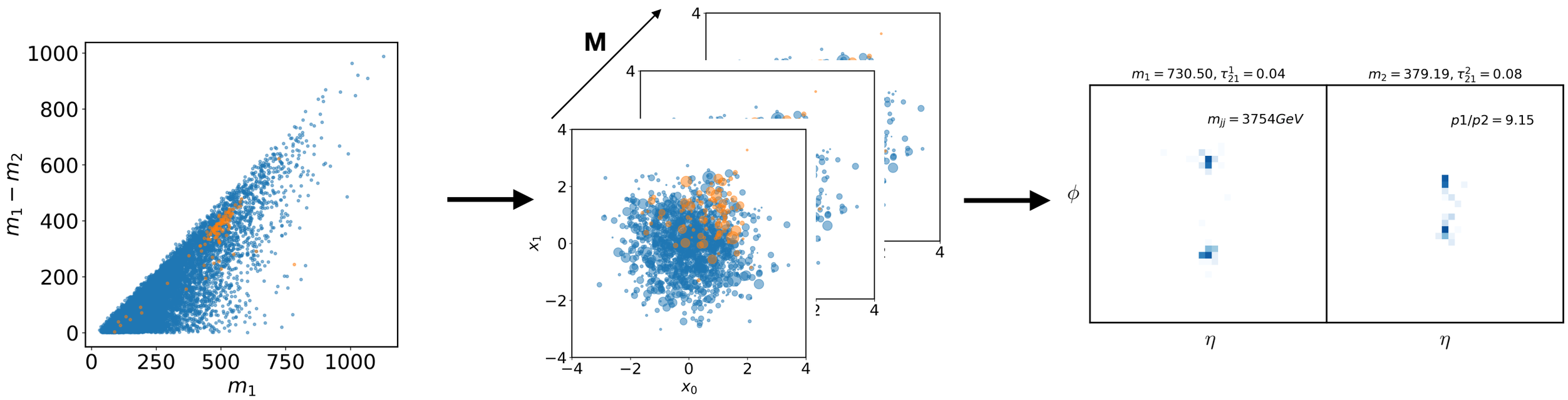
Conditional density estimation for anomaly detection

George Stein, Uros Seljak, Biwei Dai, He Jia

BERKELEY CENTER *for* COSMOLOGICAL PHYSICS

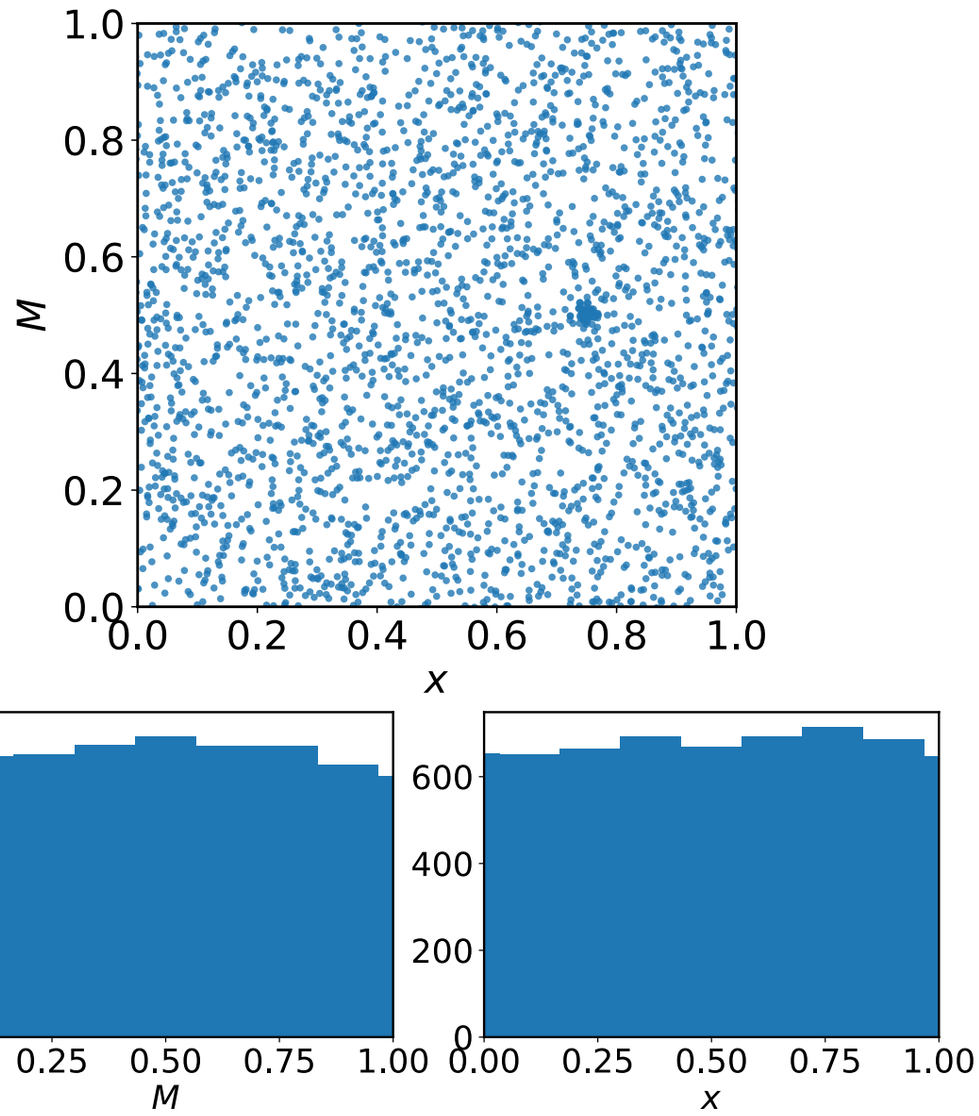


Density Estimation

Toy Example

Uniform Background

Sample Data:
10,000 events, ~1% signal
observable paramater *x*,
for event of mass *M*



Density Estimation

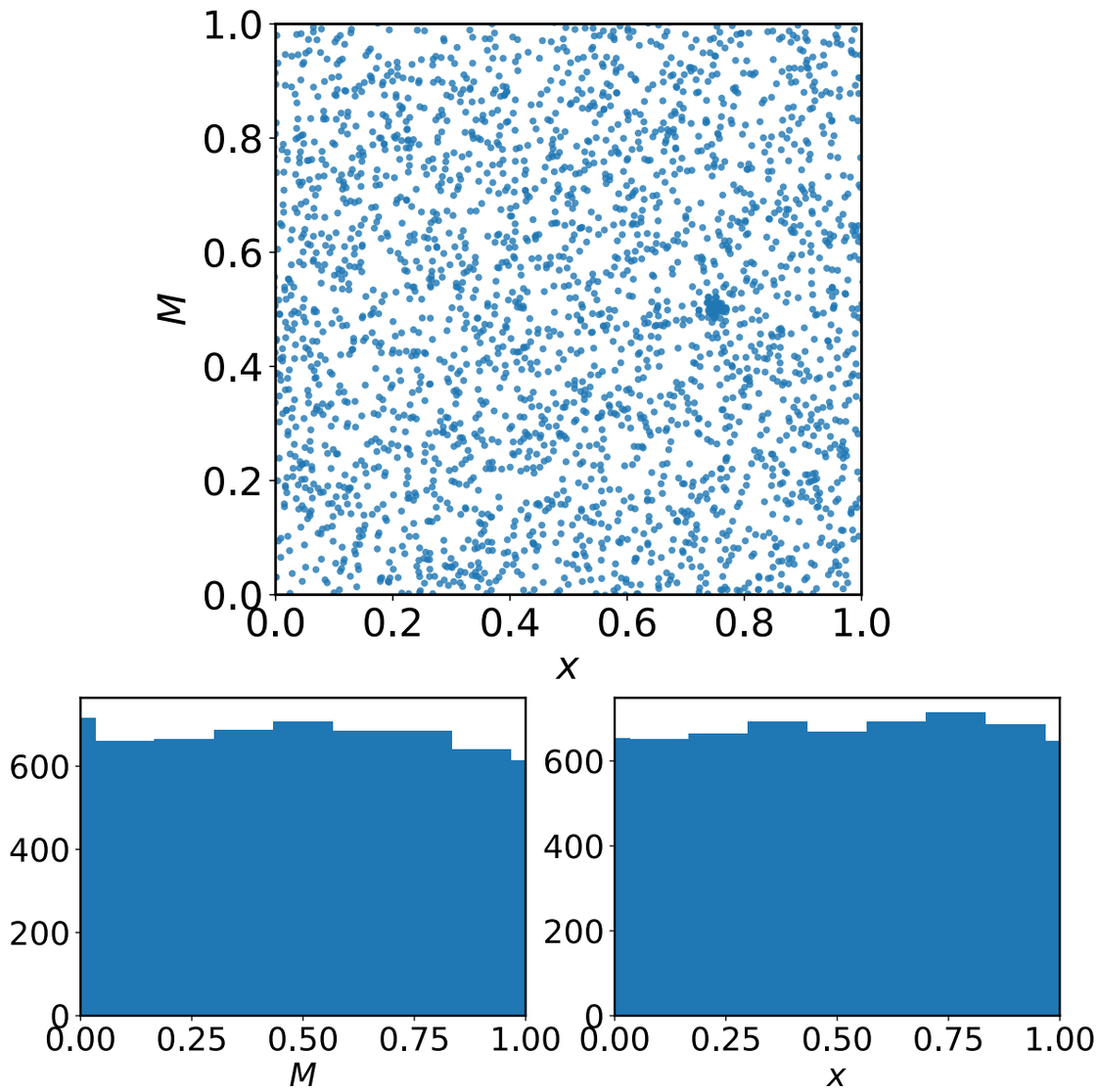
Toy Example

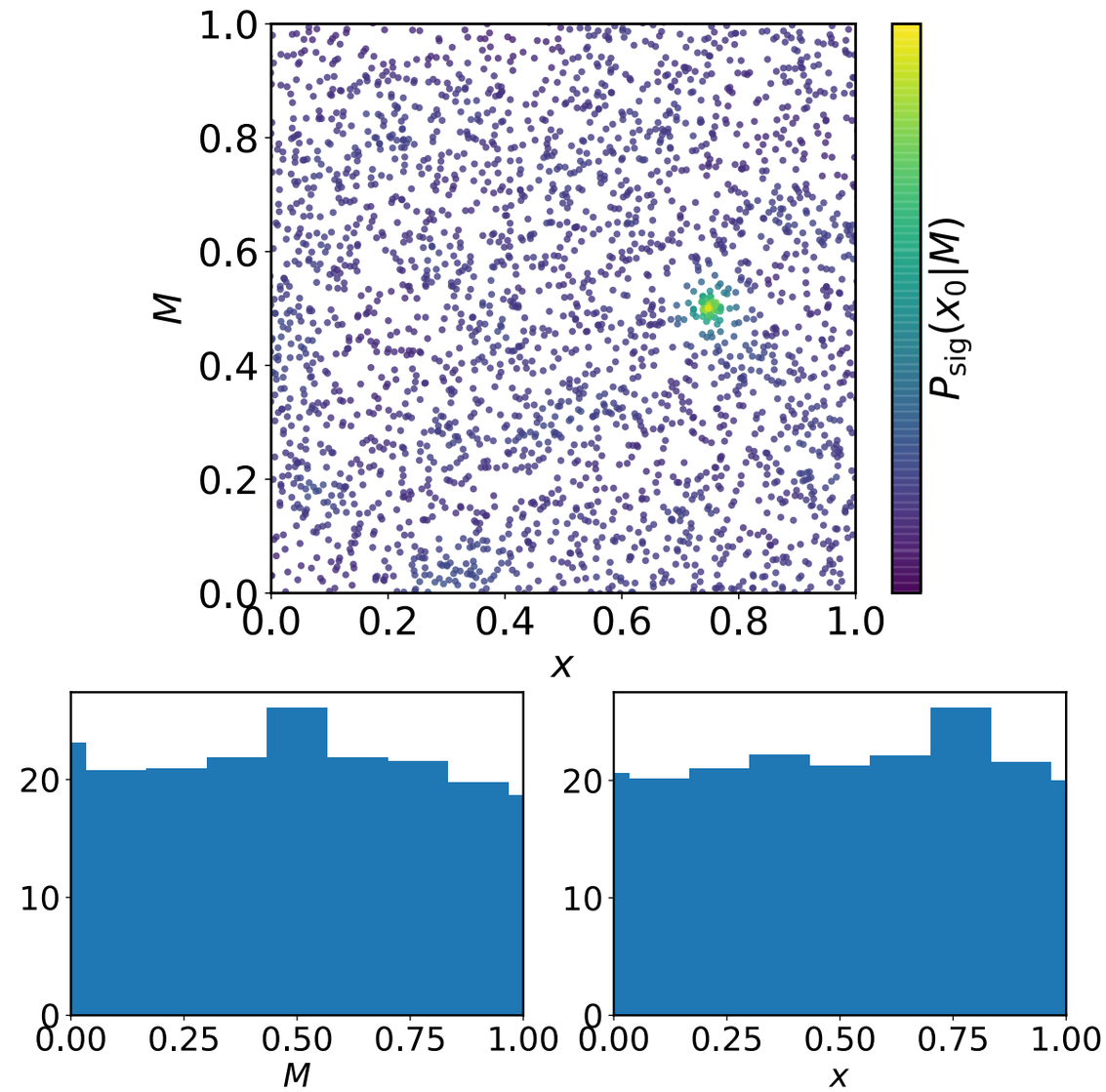
Uniform Background

Sample Data:
10,000 events, ~1% signal
observable paramater *x*,
for event of mass *M*

KNN "density",

$$P(x) \sim \frac{k}{NV_k}$$





With a uniform background, any excess is due to the signal

Density Estimation

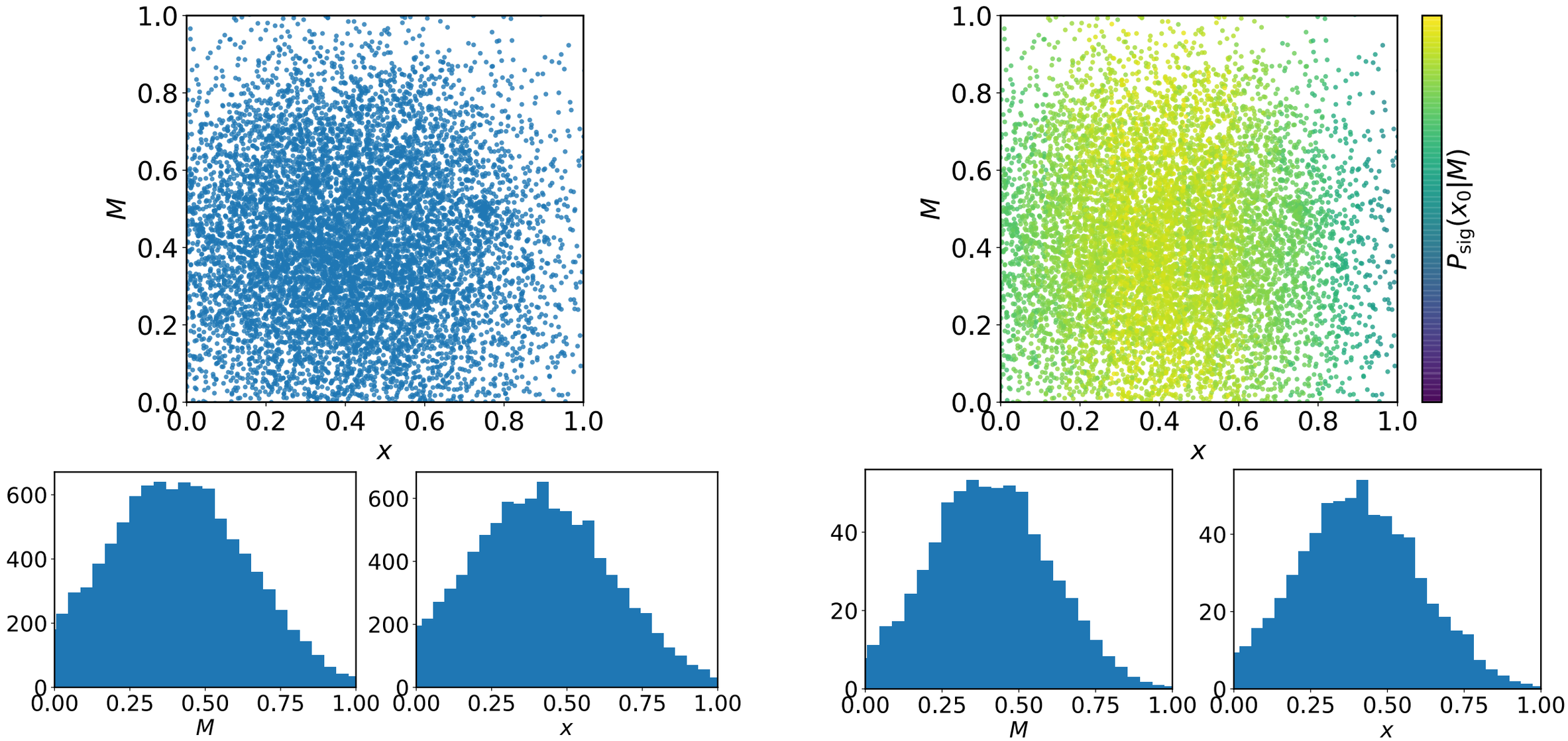
Toy Example

Nonuniform Background

Sample Data:
10,000 events, ~1% signal
observable paramater *x*,
for event of mass *M*

KNN "density",

$$P(x) \sim \frac{k}{NV_k}$$



With a nonuniform background, excess density is no longer from signal.

Instead need local overdensity P_{sig}/P_{bg}

$$P_{sig} = P(x|M) \qquad P_{bg} \sim \frac{P(x|M+\Delta) + P(x|M-\Delta)}{2}$$

Density Estimation

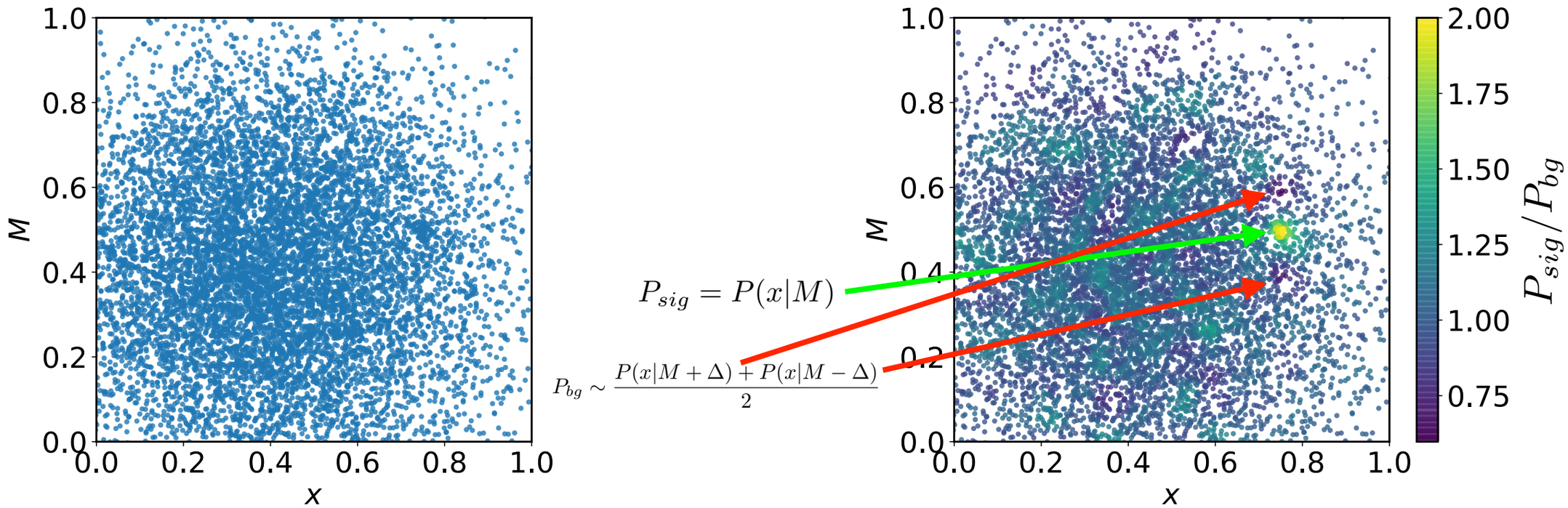
Toy Example

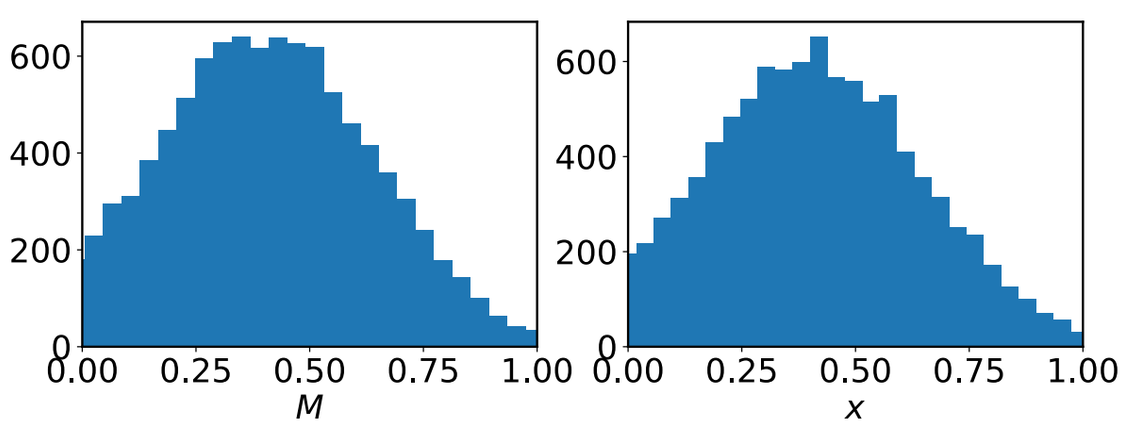
Nonuniform Background

Sample Data:
10,000 events, ~1% signal
observable paramater *x*,
for event of mass *M*

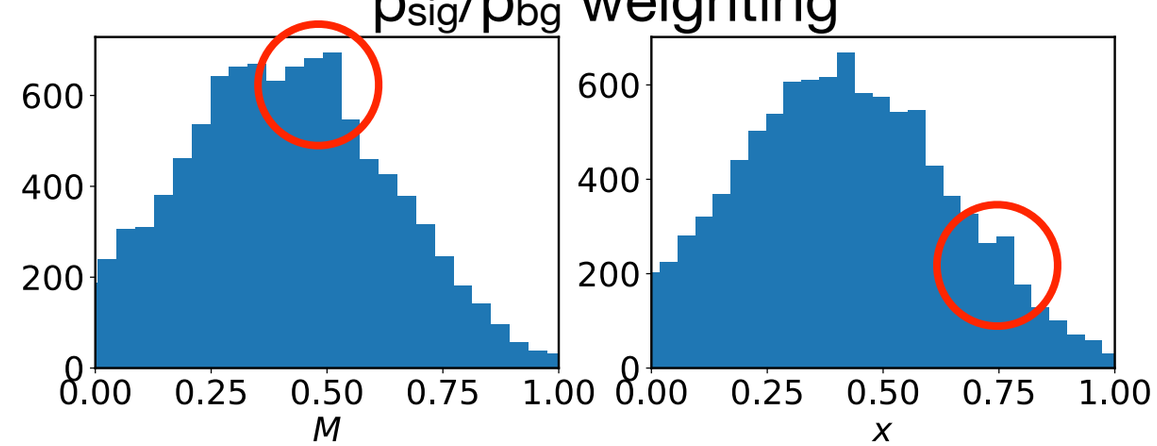
KNN "density",

$$P(x) \sim \frac{k}{NV_k}$$





p_{sig}/p_{bg} weighting



Assuming a locally smooth background, p_{sig}/p_{bg} peaks at local over densities

George Stein
BERKELEY CENTER *for* COSMOLOGICAL PHYSICS

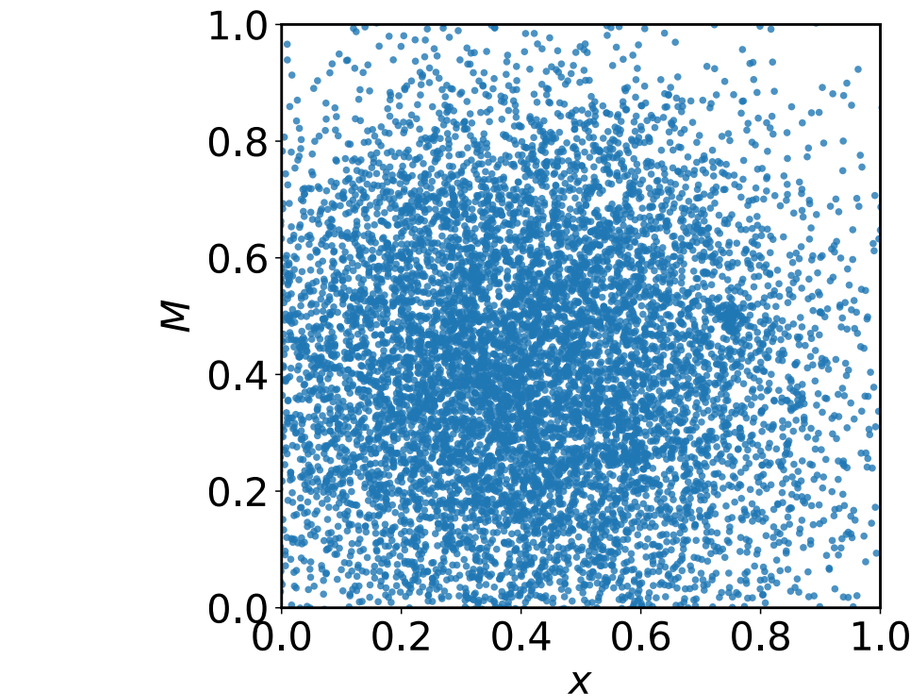
Density Estimation

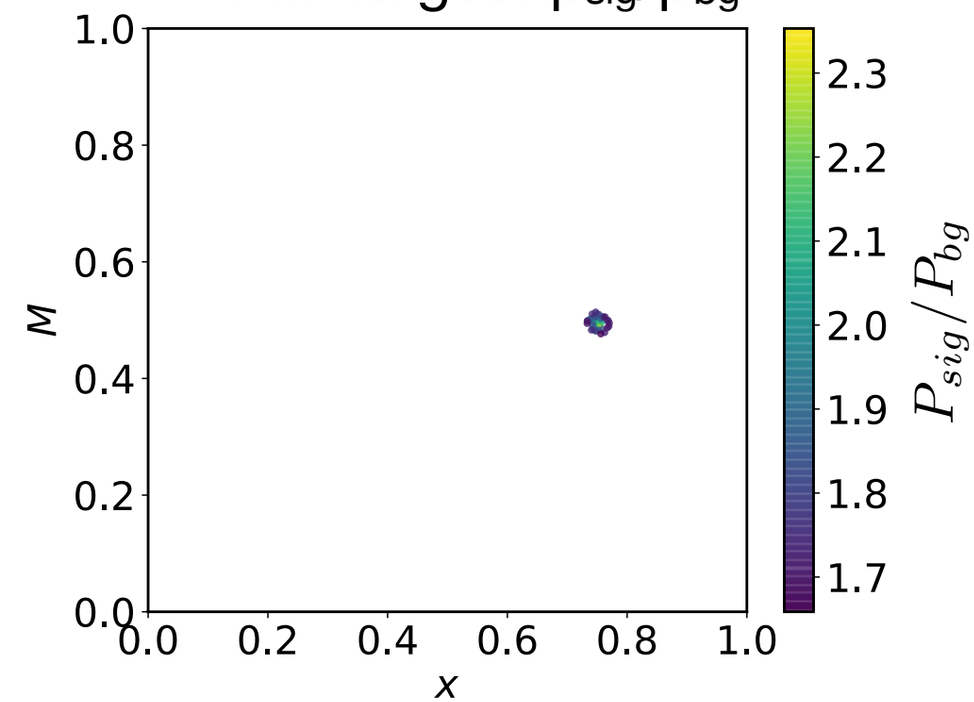
Toy Example

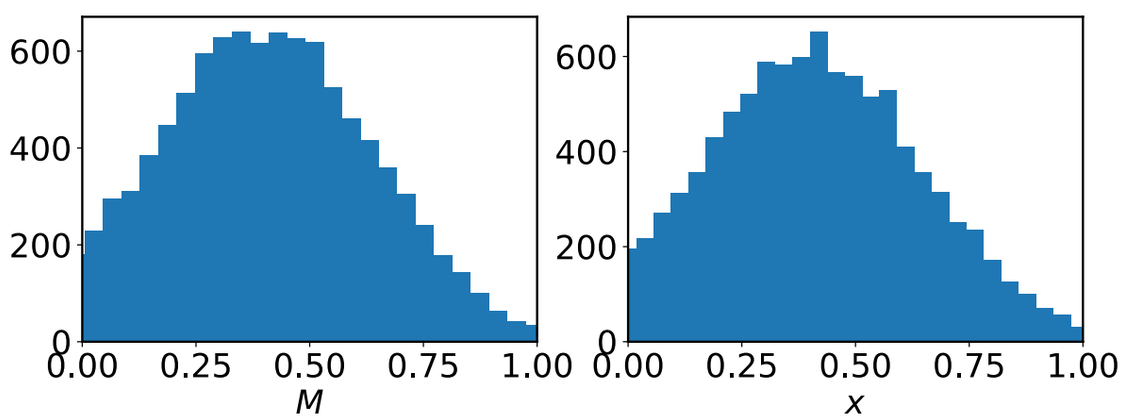
Nonuniform Background

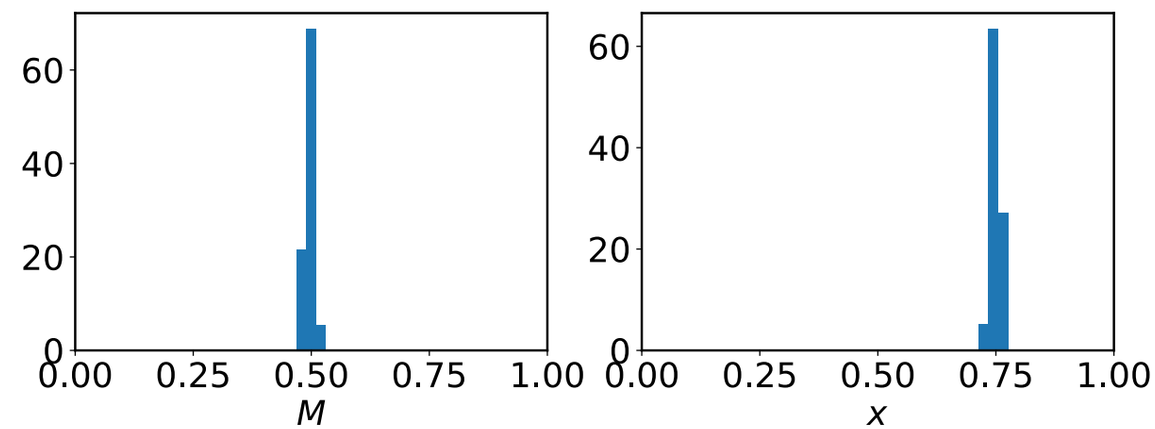
Sample Data:
10,000 events, ~1% signal
observable paramater *x*,
for event of mass *M*

Take 50 samples
with largest p_{sig}/p_{bg}









largest values of p_{sig}/p_{bg} pick out signal

KNN Density Estimation

in Original Data Space

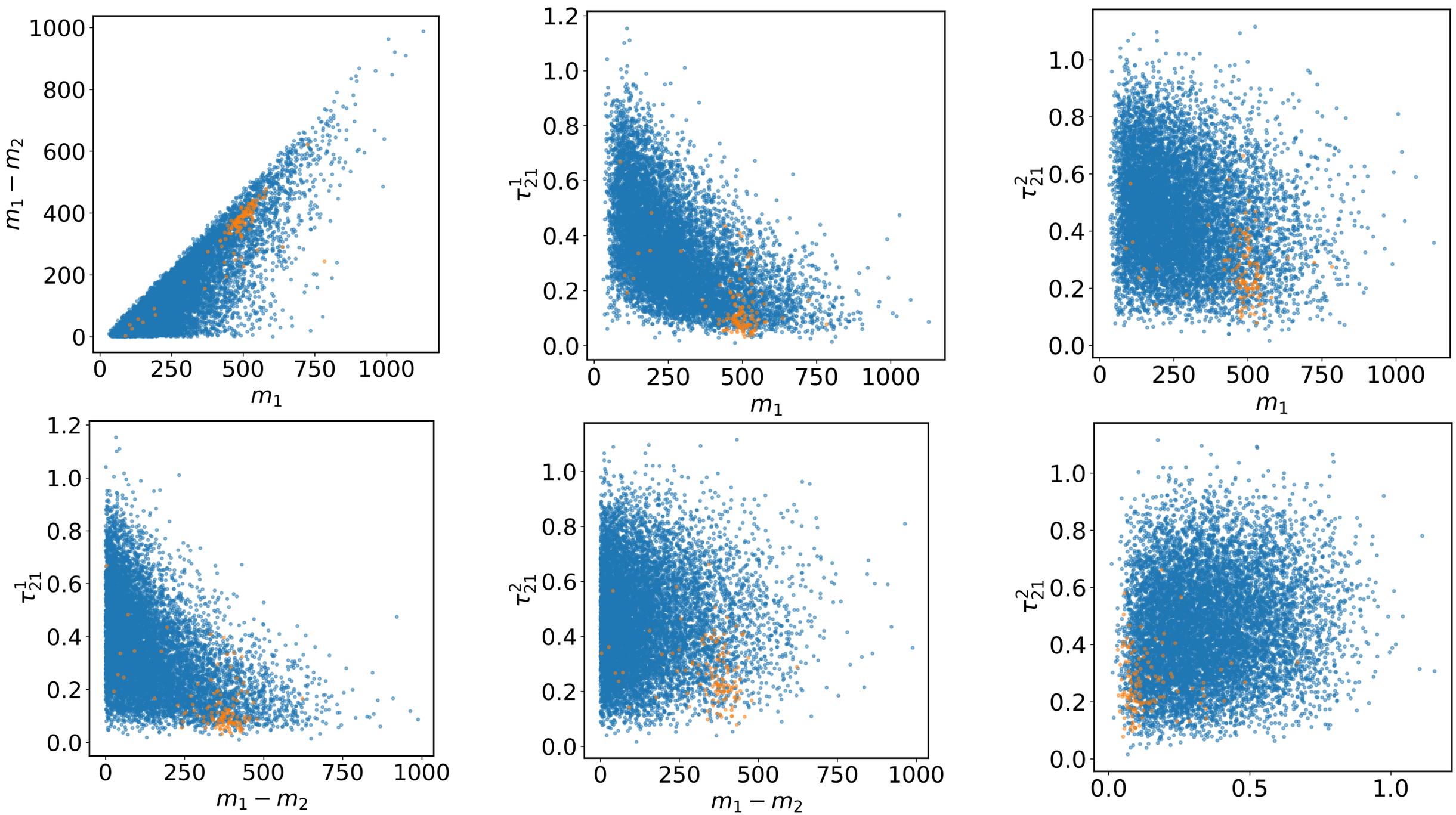
LHCO R&D example

1% signal

LHCO R&D: $m_{J1}, m_{J1} - m_{J2}, \tau_{21}^{J1}, \tau_{21}^{J2}$

pyjet with R=1 $m_{J1} > m_{J2}$ *1011.2268*

Signal at M_{JJ}=3500 GeV, m_1=500 GeV, m_1-m_2=400 GeV, ...



KNN Density Estimation
in Original Data Space
LHCO R&D example

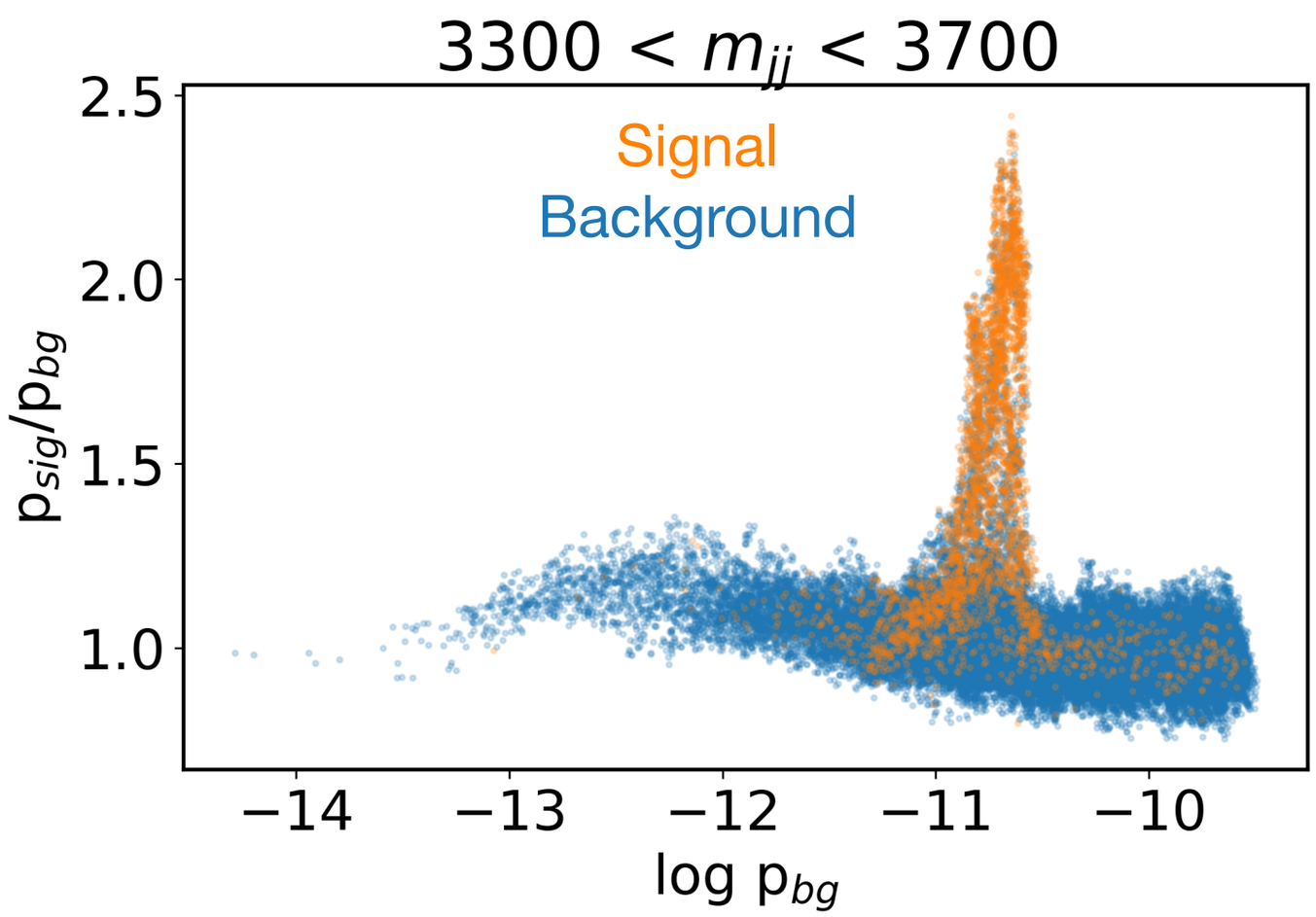
LHCO R&D: $m_{J1}, m_{J1} - m_{J2}, \tau_{21}^{J1}, \tau_{21}^{J2}$
pyjet with R=1 $m_{J1} > m_{J2}$

$$P_{sig} = P(x|M)$$

evaluate density in 3300 < M_{JJ} < 3700 GeV

$$P_{bg} \sim \frac{P(x|M+\Delta) + P(x|M-\Delta)}{2}$$

evaluate background from neighbouring sidebins
2900 < M_{JJ} < 3300 GeV, 3700 < M_{JJ} < 4100 GeV



signal events
picked out

but highly contaminated
with background

Density Estimation

KNN density in original data space is not ideal

$$P(x) \sim \frac{k}{NV_k}$$

- Not a real density
- Noisy
- Low density regions have too large of estimated density (heavy tails)
- Not smooth

Transform data to better distribution

Density Estimation Through Conditional Iterative Gaussianization

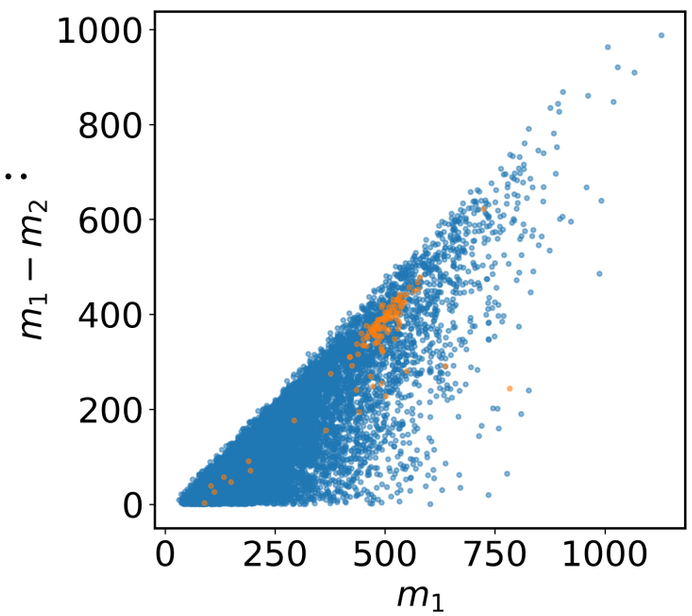
(working name)

LHCO R&D: $m_{J1}, m_{J1} - m_{J2}, \tau_{21}^{J1}, \tau_{21}^{J2}$

pyjet with R=1 $m_{J1} > m_{J2}$

1. Input Data *x*

$$p(x|M) = \ldots$$



Many methods for density estimation

"Gaussianization"

NIPS2000_1856

1602.00229

NN based:

Normalizing Flows

1505.05770

NAF

1804.00779

MAF

1705.07057

hep 2001.04990

● *size proportional determinant of Jacobian*

Density Estimation Through Conditional Iterative Gaussianization

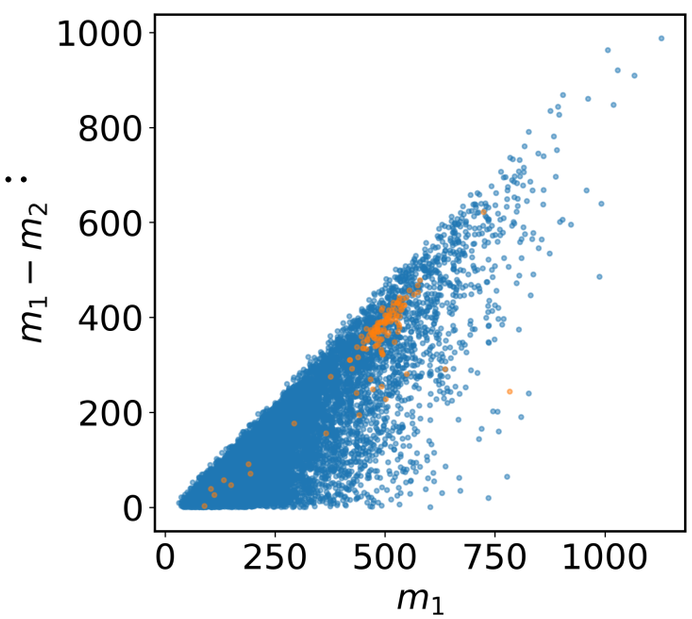
(working name)

LHCO R&D: $m_{J1}, m_{J1} - m_{J2}, \tau_{21}^{J1}, \tau_{21}^{J2}$

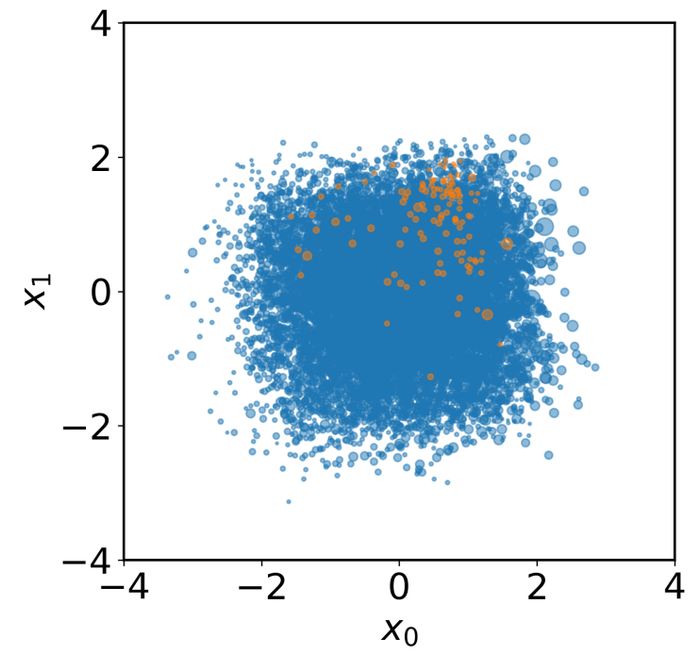
pyjet with R=1 $m_{J1} > m_{J2}$

$$p(x|M) = \ldots$$

1. Input Data *x*



2. Whiten Data



Many methods for density estimation

"Gaussianization"

NIPS2000_1856

1602.00229

NN based:

Normalizing Flows

1505.05770

NAF

1804.00779

MAF

1705.07057

hep 2001.04990

- *size proportional determinant of Jacobian*

Density Estimation Through Conditional Iterative Gaussianization

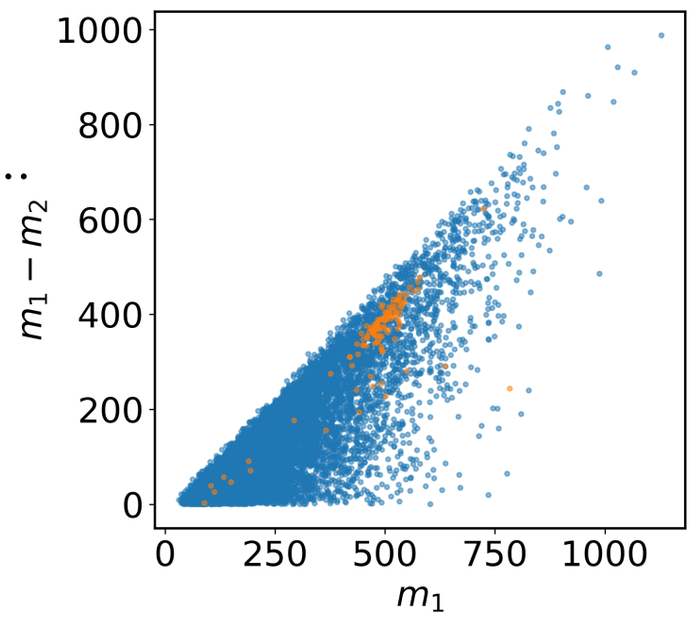
(working name)

LHCO R&D: $m_{J1}, m_{J1} - m_{J2}, \tau_{21}^{J1}, \tau_{21}^{J2}$

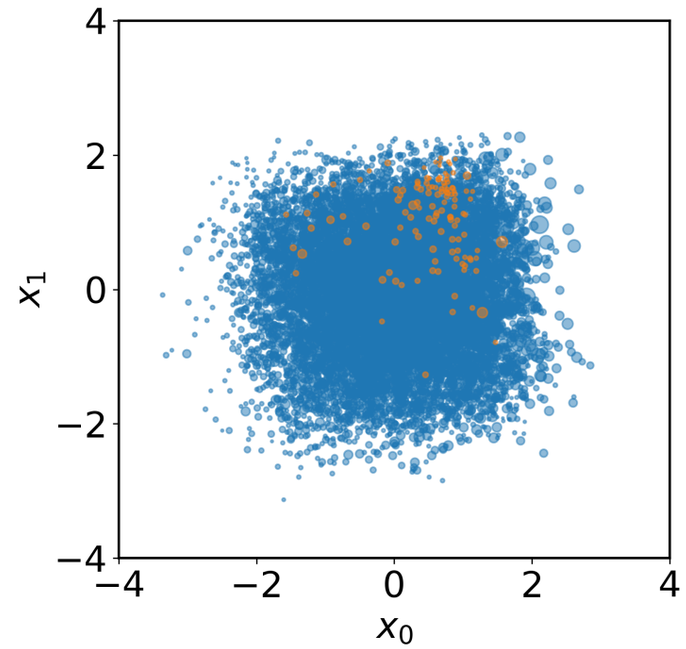
pyjet with R=1 $m_{J1} > m_{J2}$

$p(x|M) = \ldots$

1. Input Data *x*

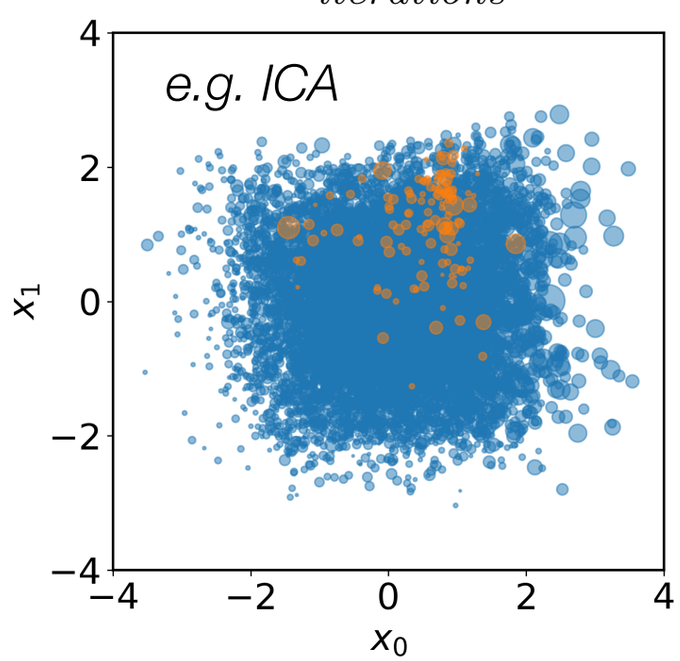


2. Whiten Data



3. Uncorrelate Features

$n_{iterations}$



Many methods for density estimation

"Gaussianization"

NIPS2000_1856

1602.00229

NN based:

Normalizing Flows

1505.05770

NAF

1804.00779

MAF

1705.07057

hep 2001.04990

- *size proportional determinant of Jacobian*

Density Estimation Through Conditional Iterative Gaussianization

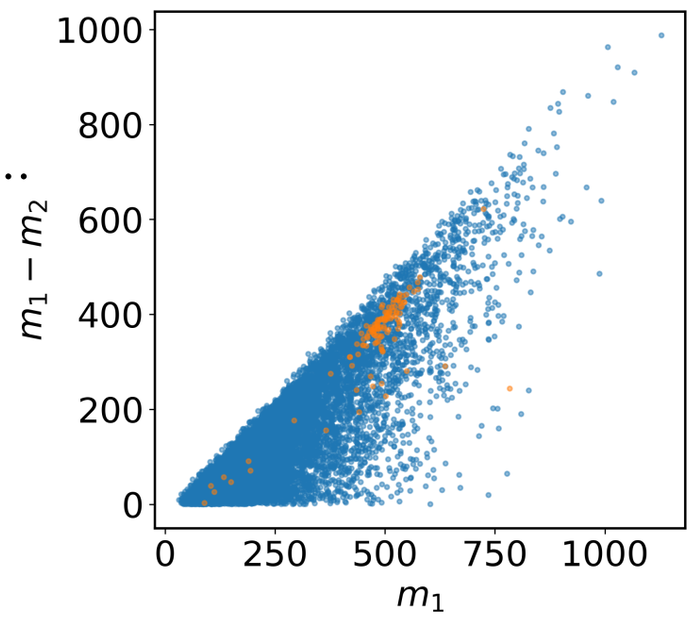
(working name)

LHCO R&D: $m_{J1}, m_{J1} - m_{J2}, \tau_{21}^{J1}, \tau_{21}^{J2}$

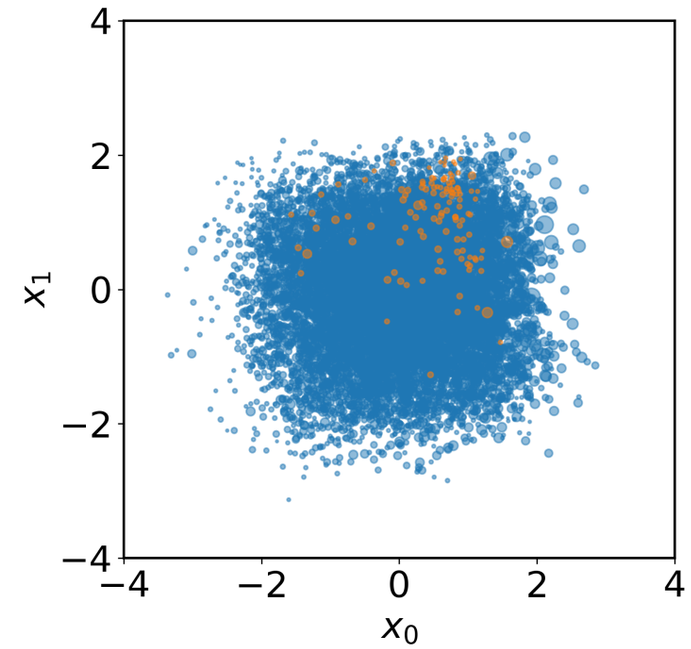
pyjet with R=1 $m_{J1} > m_{J2}$

$p(x|M) = \ldots$

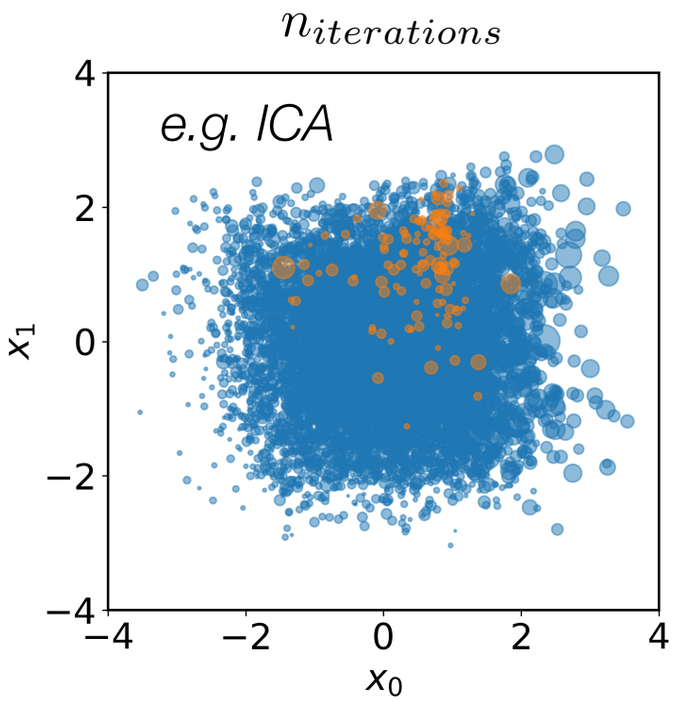
1. Input Data *x*



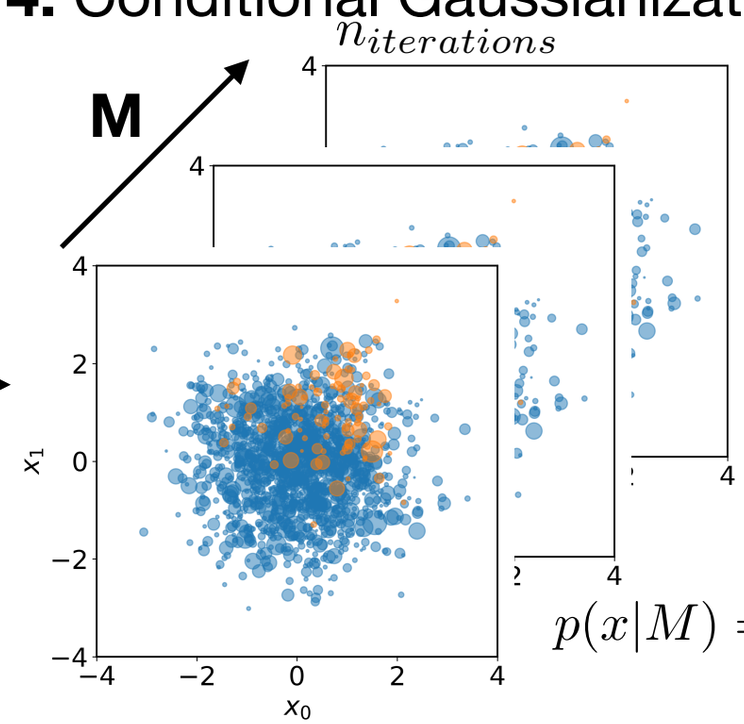
2. Whiten Data



3. Uncorrelate Features



4. Conditional Gaussianization



$$p(x|M) = p_y(f(x|M)) \left| \frac{df(x|M)}{dx} \right|$$

Not limited to Gaussian

Many methods for density estimation

"Gaussianization"

NIPS2000_1856

1602.00229

NN based:

Normalizing Flows

1505.05770

NAF

1804.00779

MAF

1705.07057

hep 2001.04990

- *size proportional determinant of Jacobian*

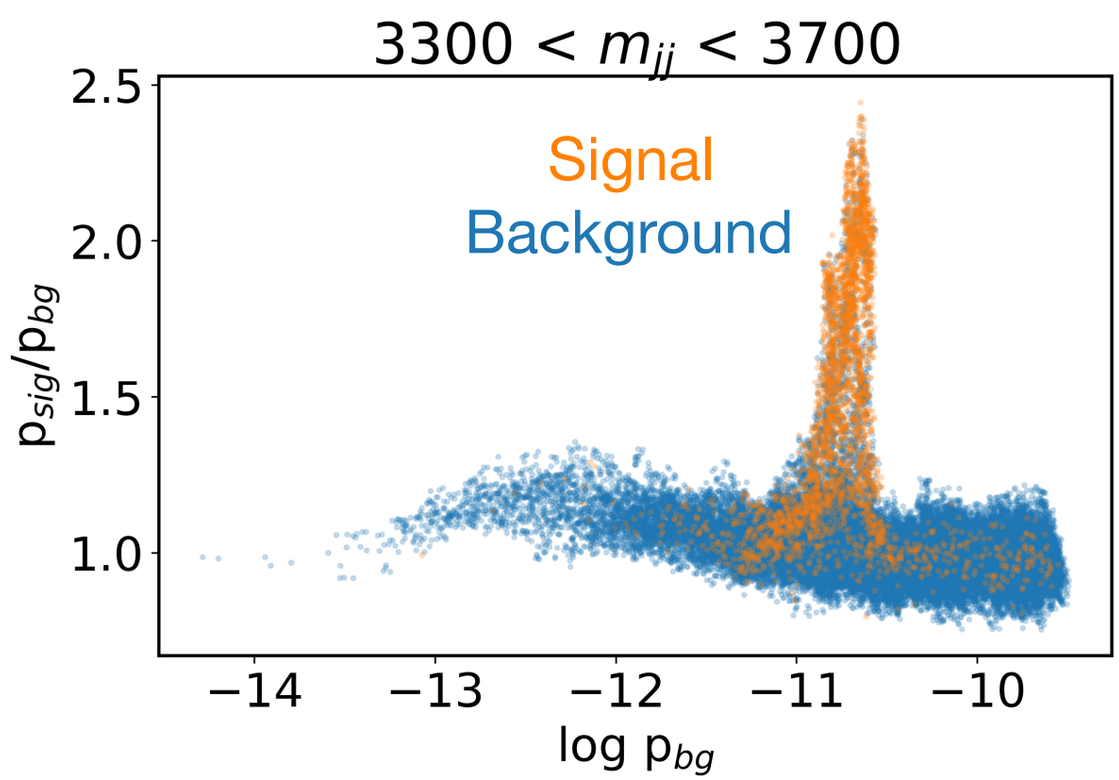
George Stein

BERKELEY CENTER *for* COSMOLOGICAL PHYSICS

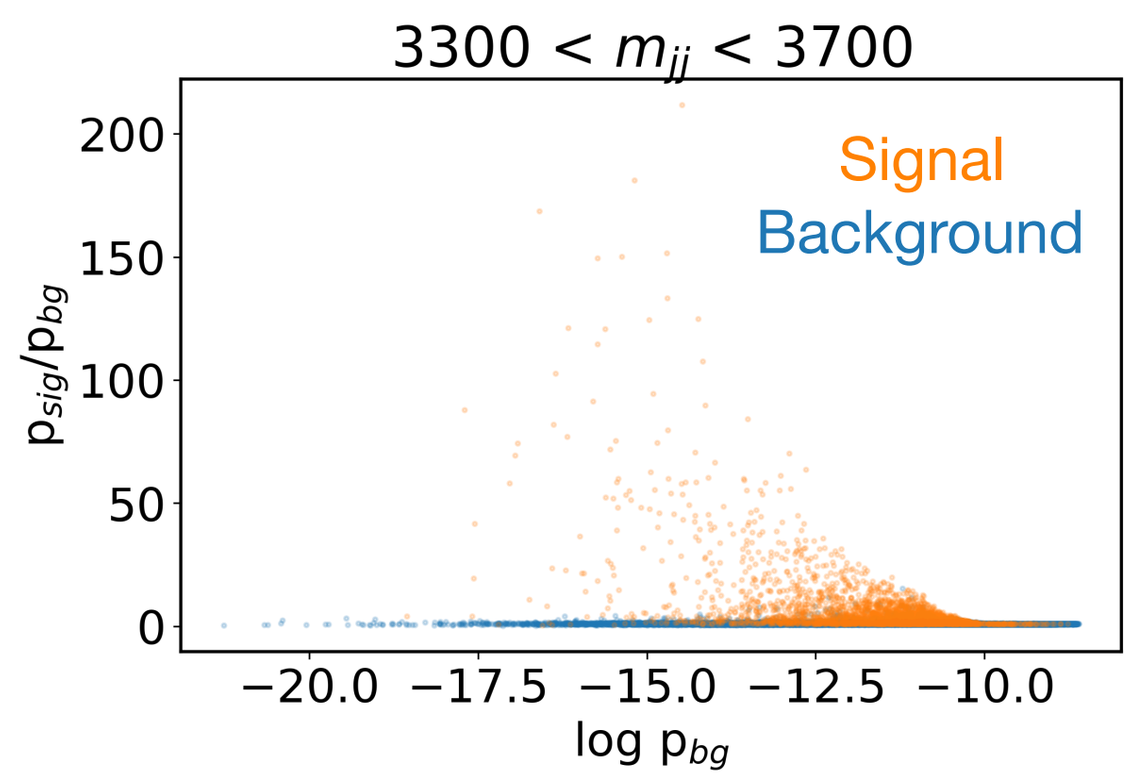
Full Conditional PDF of data
$P(\mathbf{x} \mid M_{JJ})$

No a priori knowledge of signal region required

KNN in original data space



Our Method



BlackBox 1

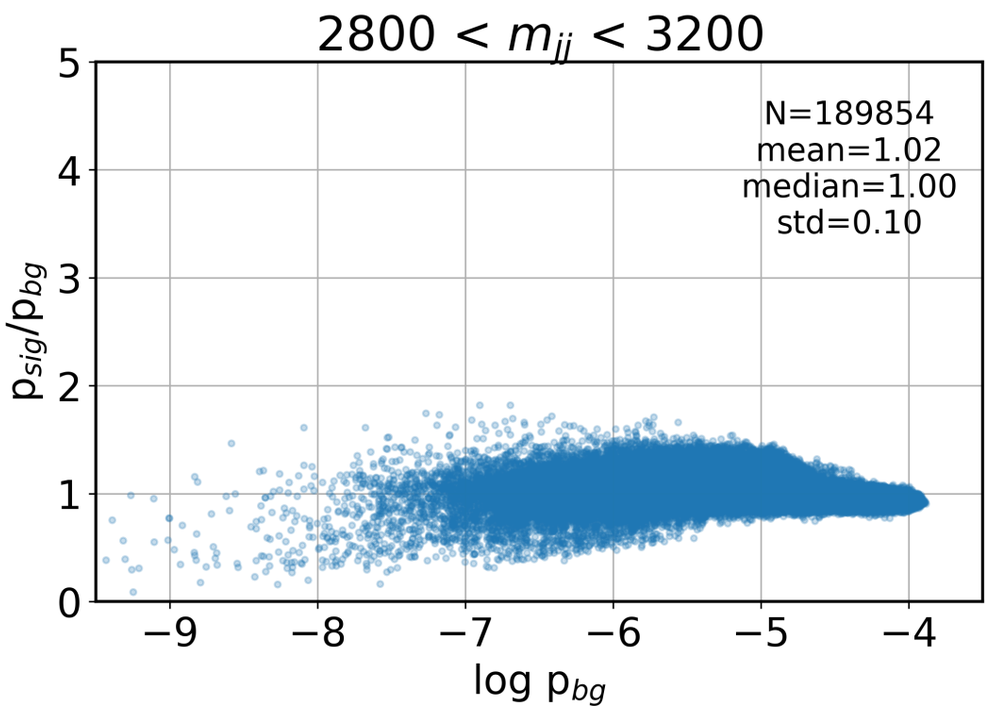
LHCO R&D: $m_{J1}, m_{J1} - m_{J2}, \tau_{21}^{J1}, \tau_{21}^{J2}$

pyjet with R=1 $m_{J1} > m_{J2}$

BlackBox 1

LHCO R&D: $m_{J1}, m_{J1} - m_{J2}, \tau_{21}^{J1}, \tau_{21}^{J2}$

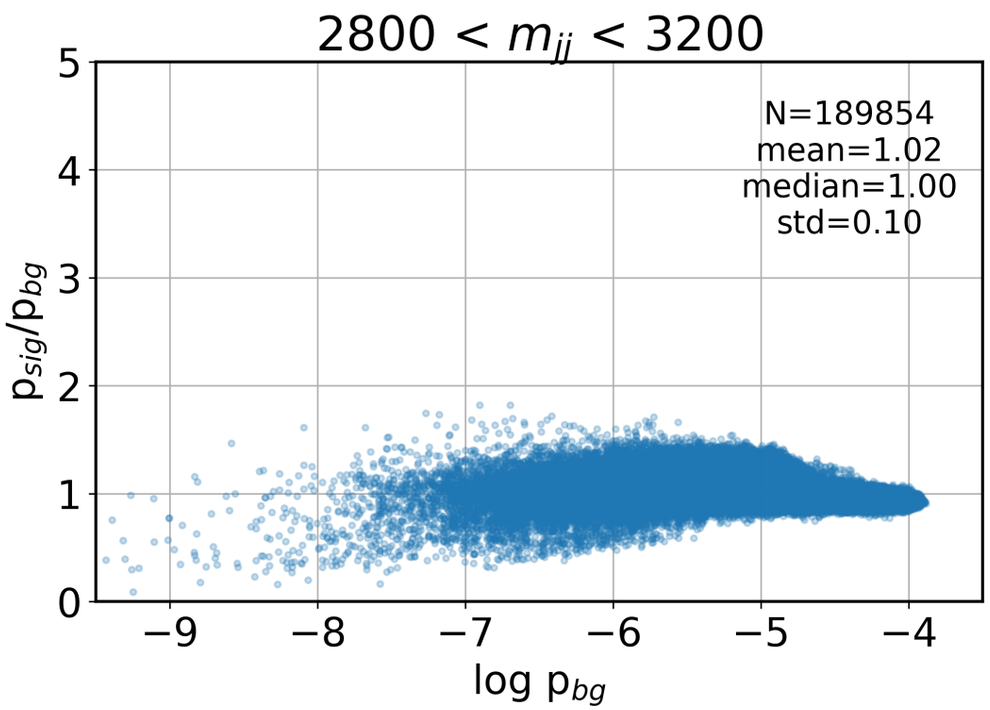
pyjet with R=1 $m_{J1} > m_{J2}$

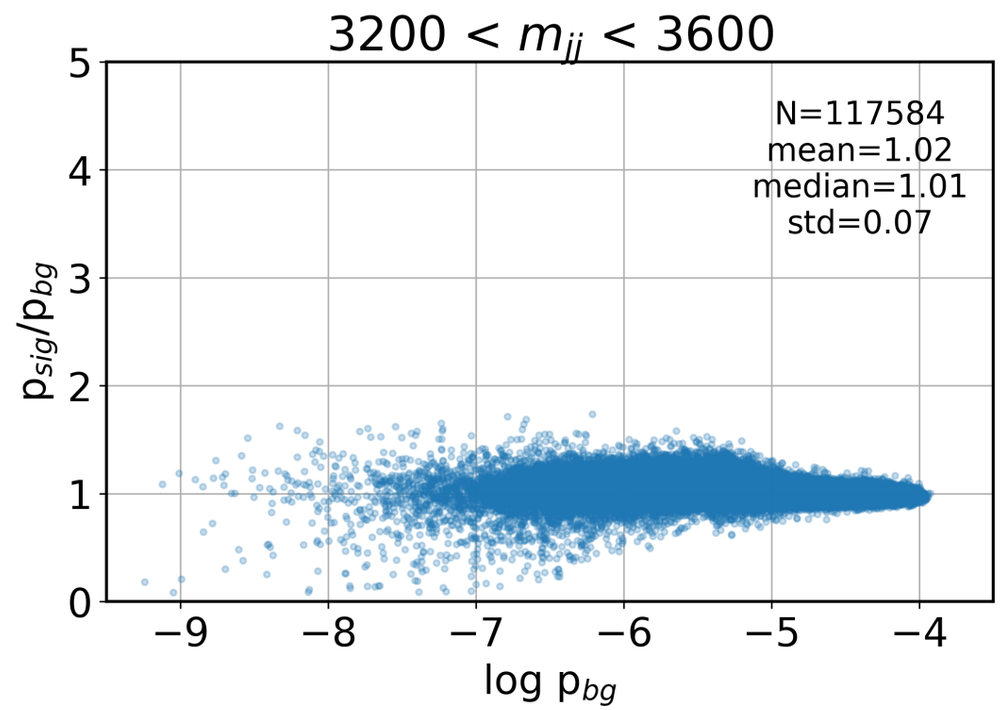


BlackBox 1

LHCO R&D: $m_{J1}, m_{J1} - m_{J2}, \tau_{21}^{J1}, \tau_{21}^{J2}$

pyjet with R=1 $m_{J1} > m_{J2}$

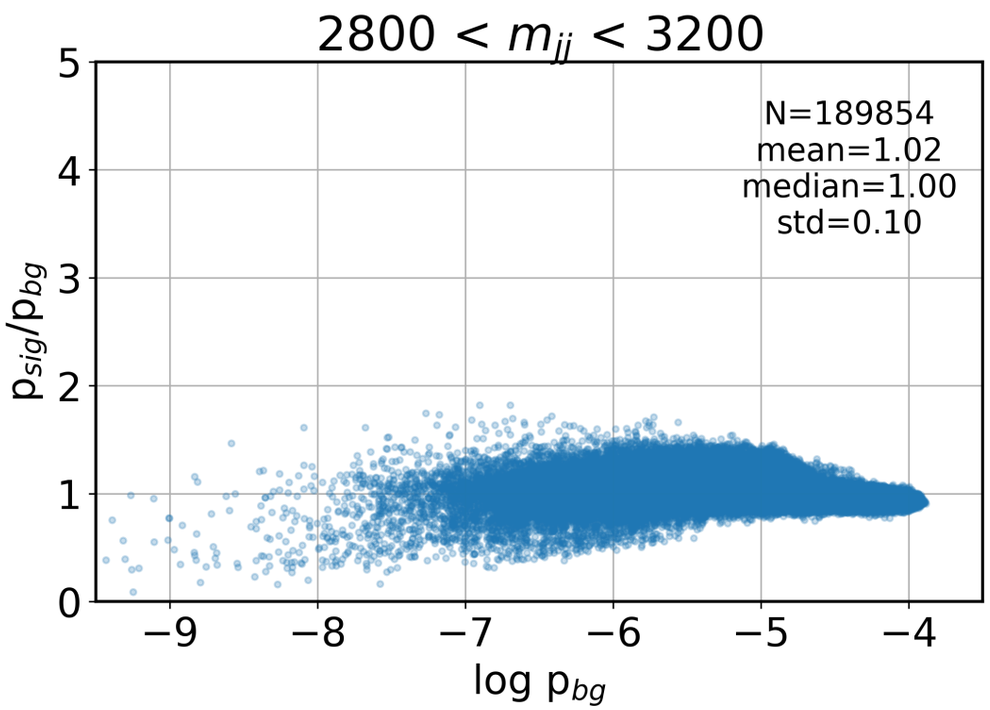


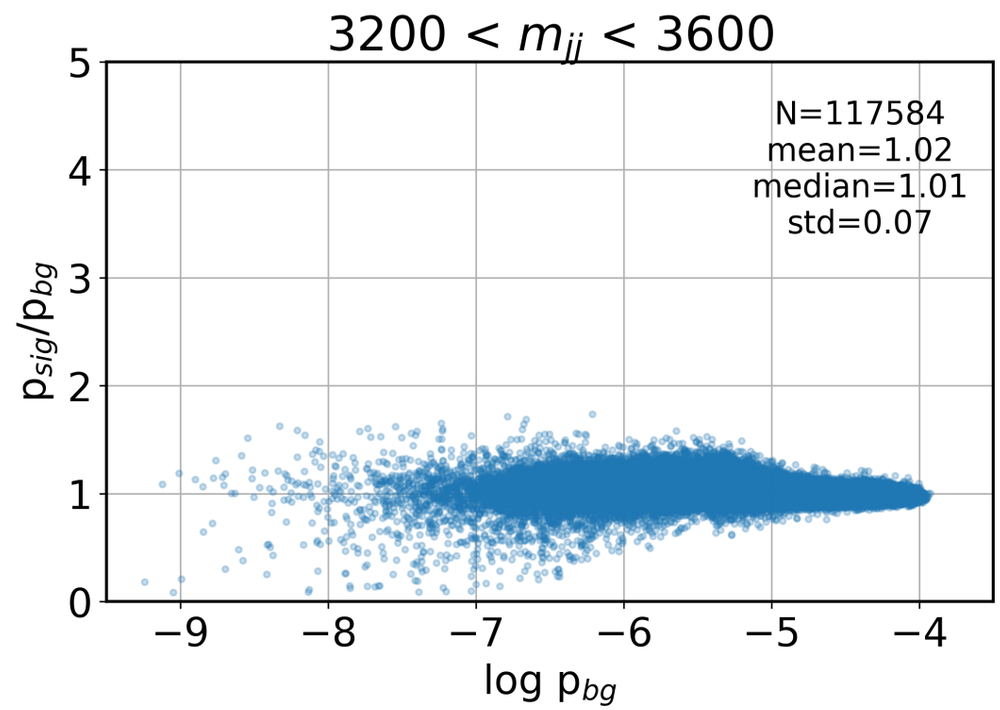


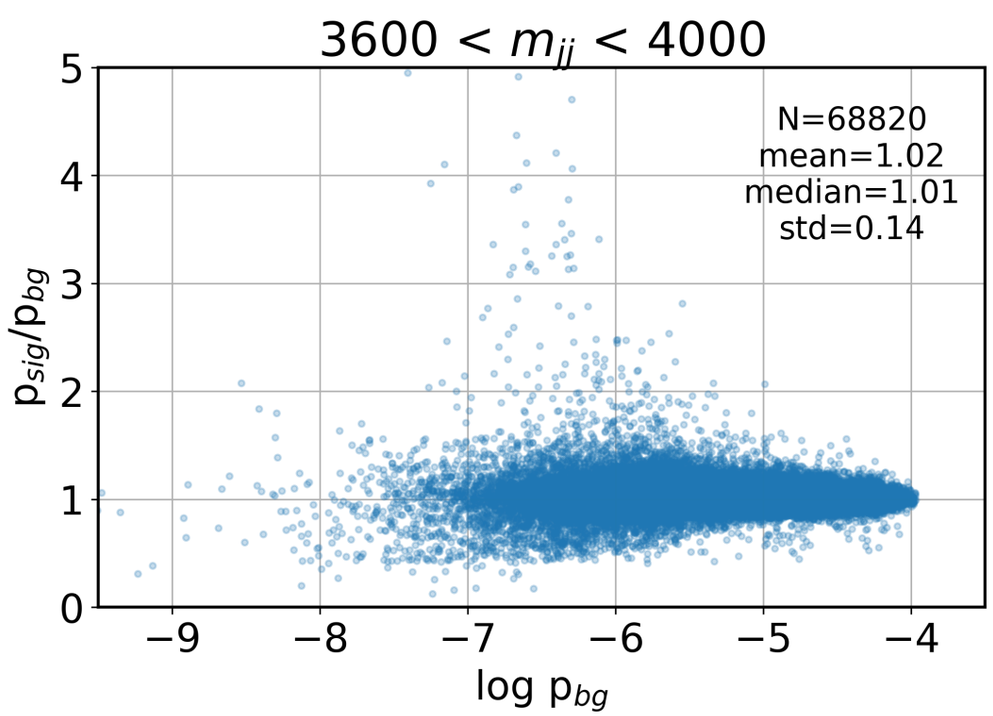
BlackBox 1

LHCO R&D: $m_{J1}, m_{J1} - m_{J2}, \tau_{21}^{J1}, \tau_{21}^{J2}$

pyjet with R=1 $m_{J1} > m_{J2}$



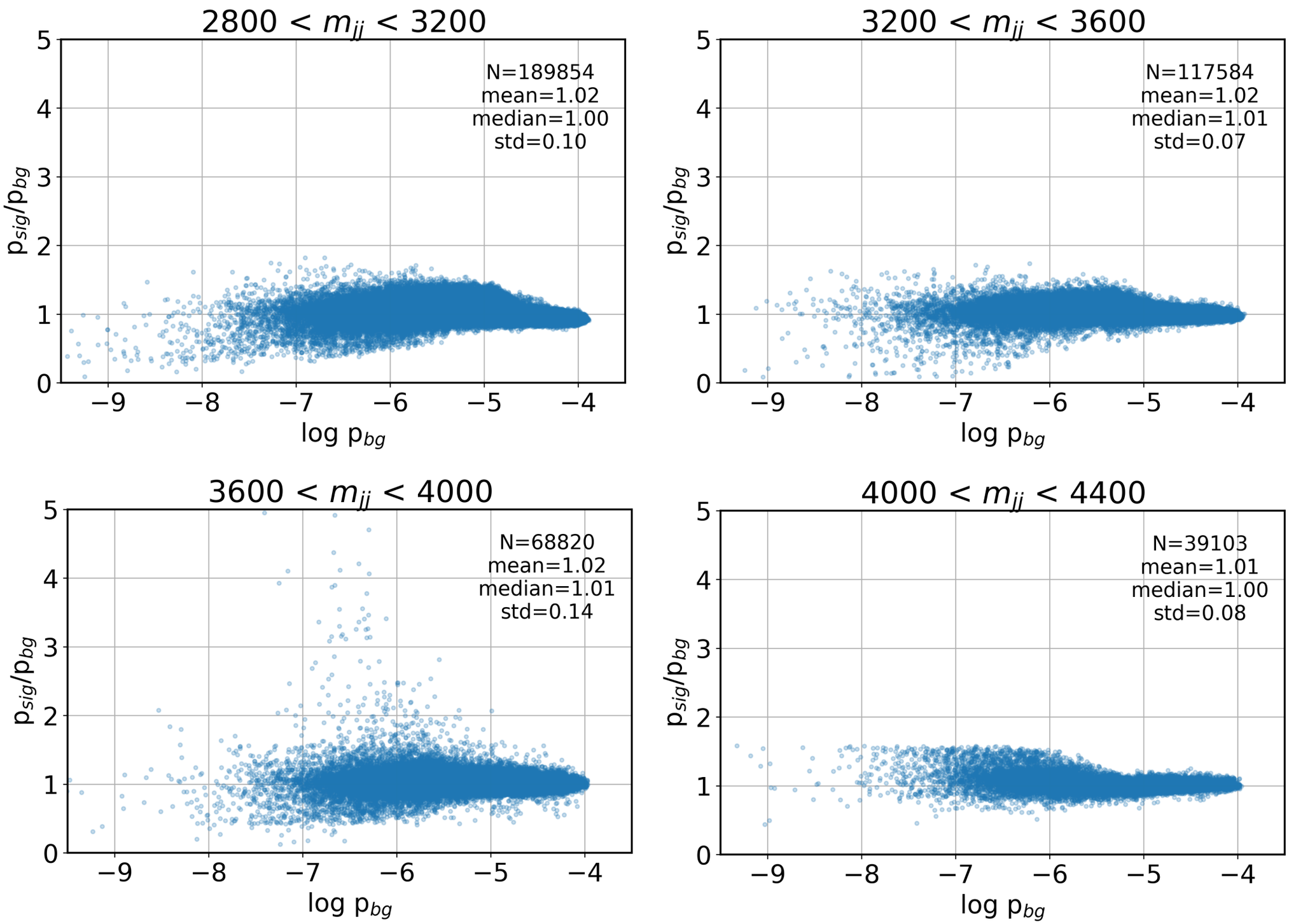




BlackBox 1

LHCO R&D: $m_{J1}, m_{J1} - m_{J2}, \tau_{21}^{J1}, \tau_{21}^{J2}$

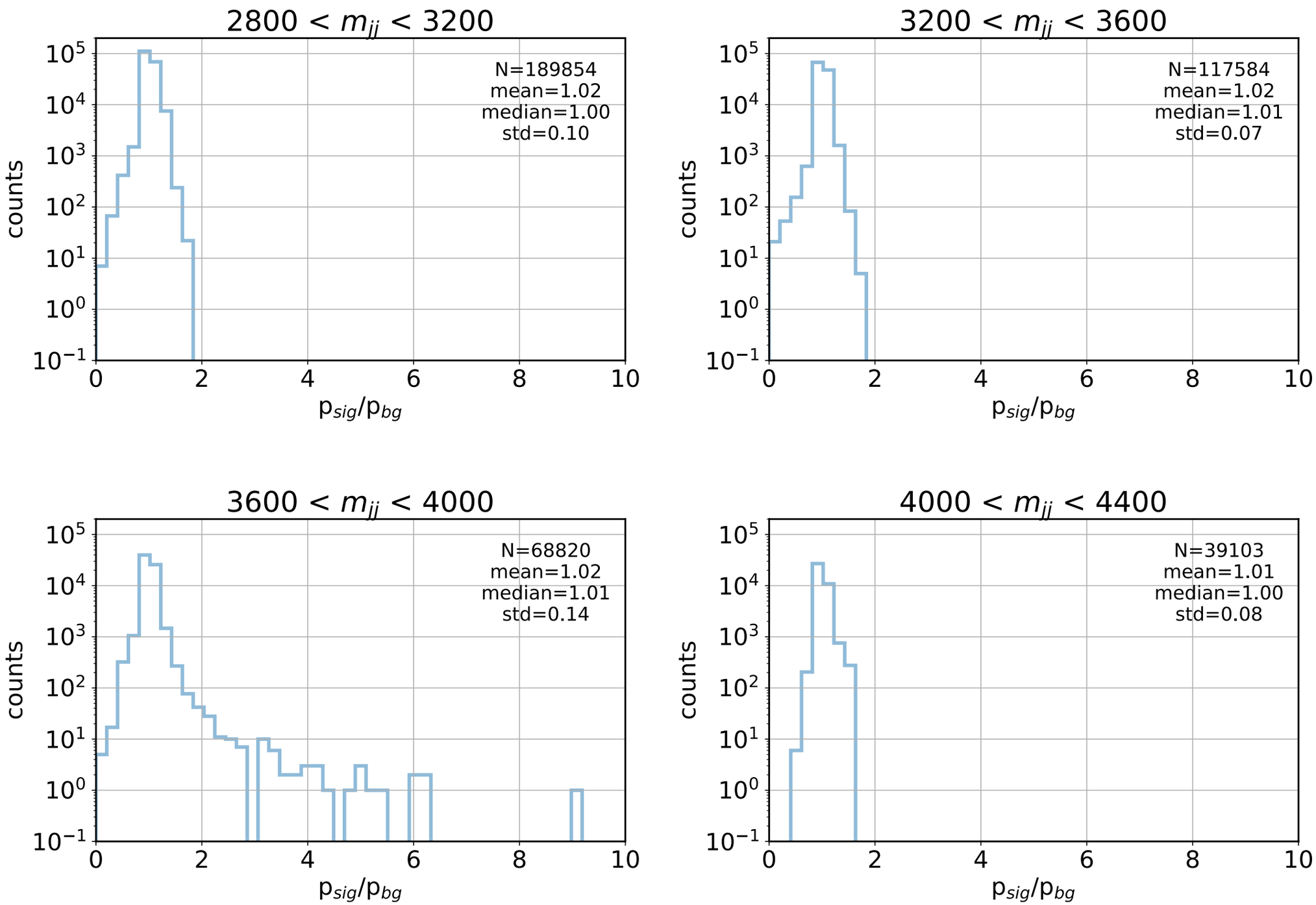
pyjet with R=1 $m_{J1} > m_{J2}$



BlackBox 1

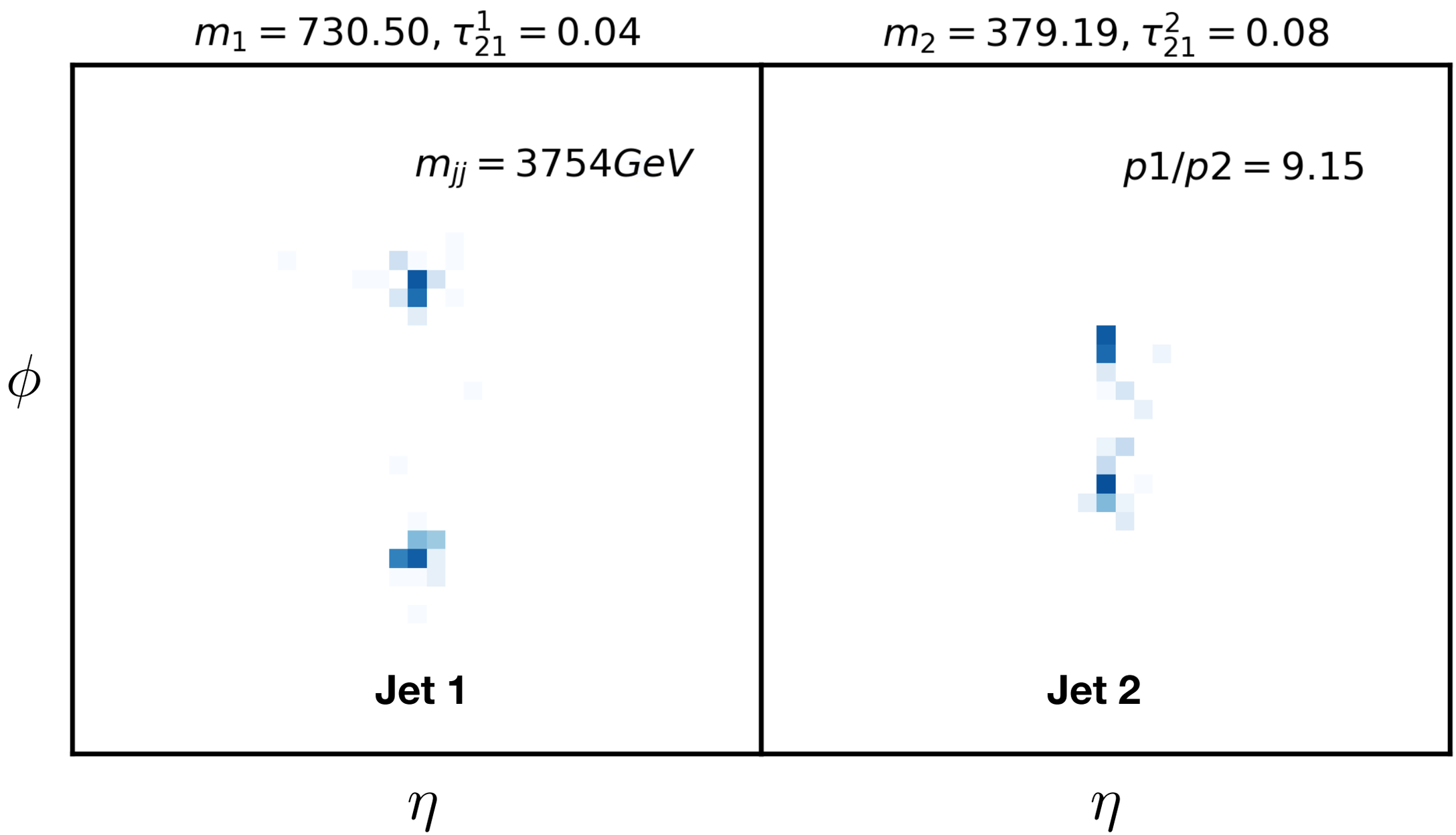
LHCO R&D: $m_{J1}, m_{J1} - m_{J2}, \tau_{21}^{J1}, \tau_{21}^{J2}$

pyjet with R=1 $m_{J1} > m_{J2}$



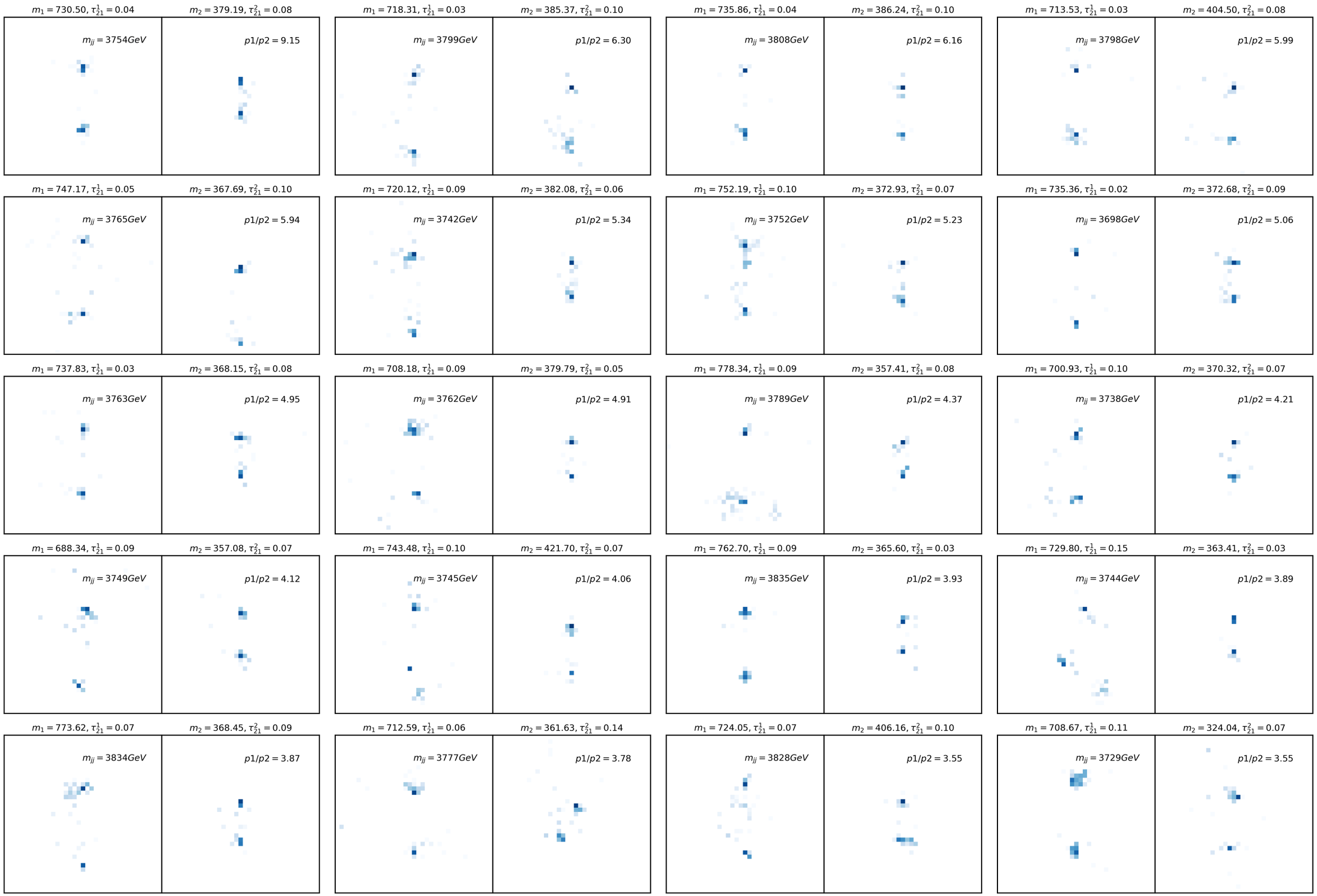
BlackBox 1

Event with largest value of p_{sig}/p_{bg}



BlackBox 1

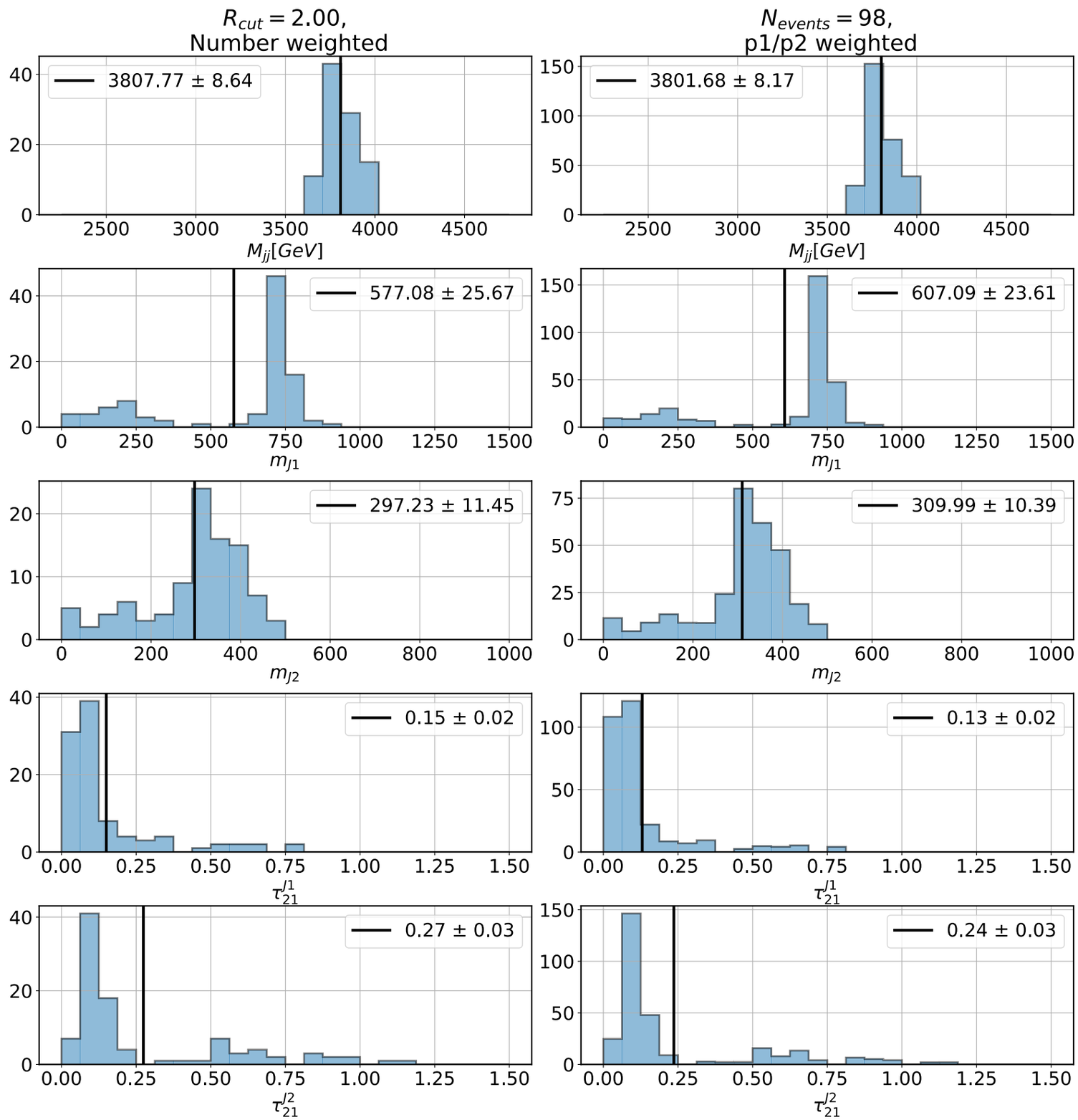
20 events with largest values of p_{sig}/p_{bg}



BlackBox 1

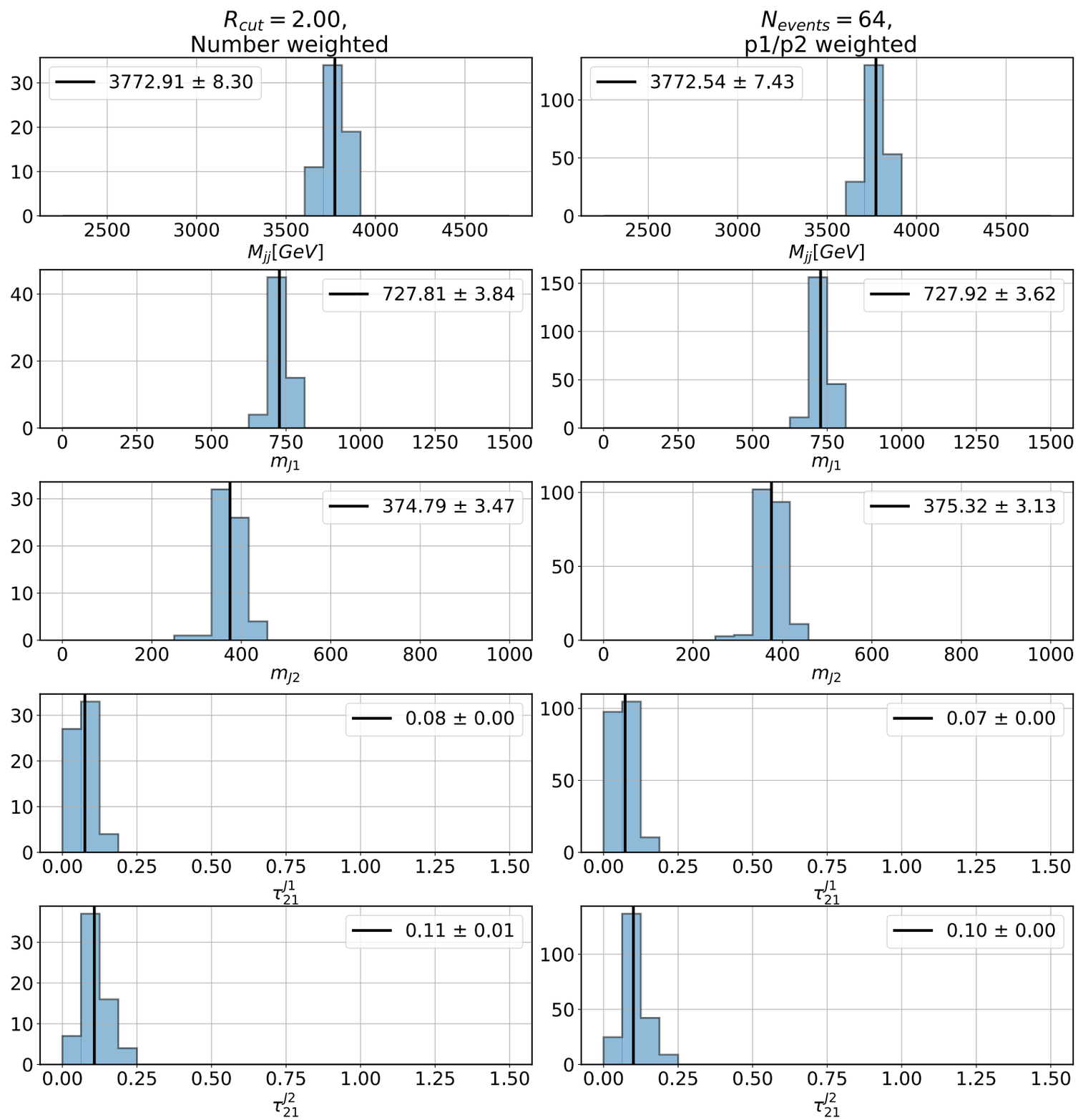
First look at summary statistics of events that survive cut

$$R = \frac{p_{sig}}{p_{bg}}$$



BlackBox 1 Cleaned look at summary statistics of events that survive cut

$$R = \frac{p_{sig}}{p_{bg}}$$



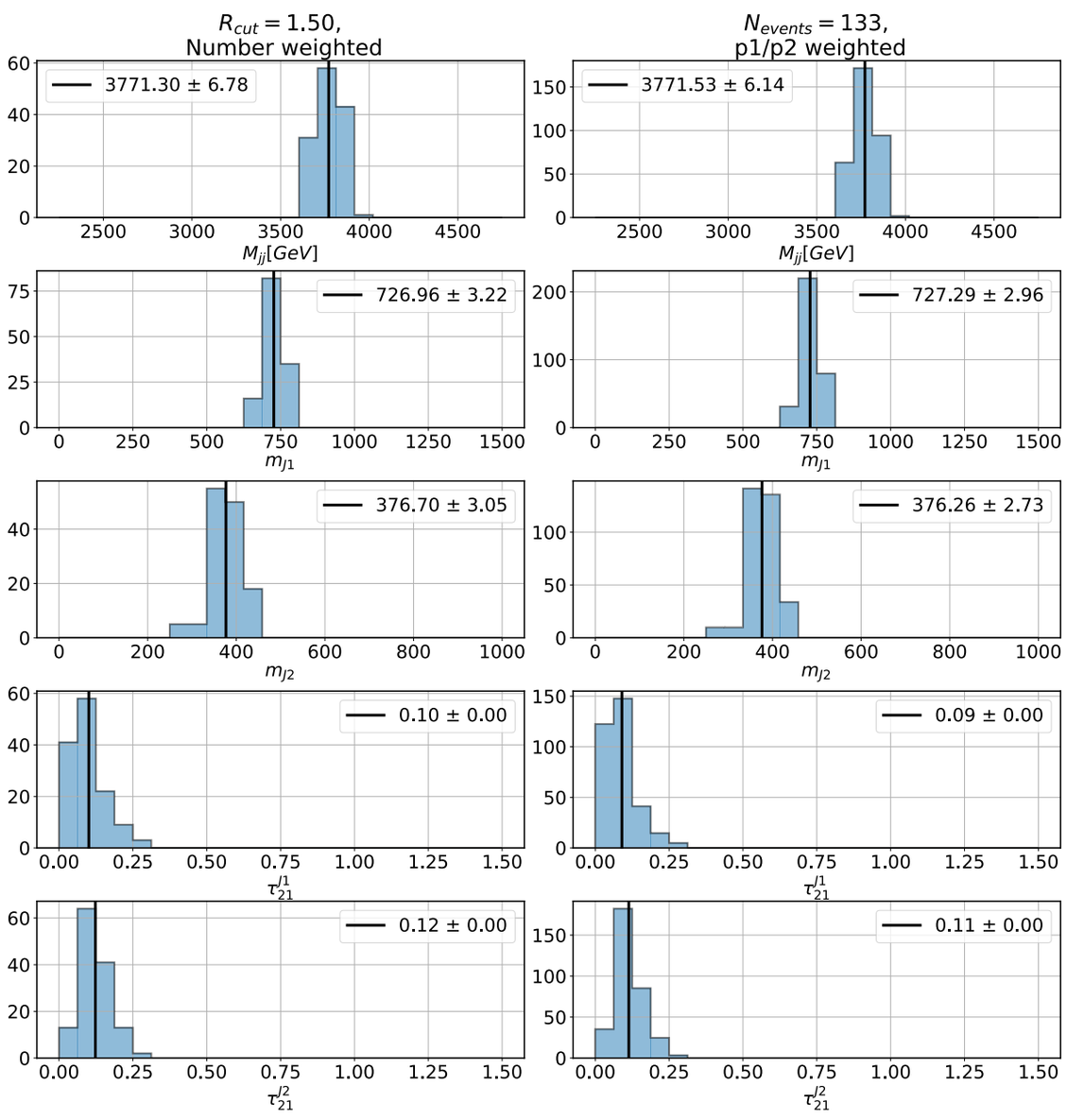
Cut outside of range:
logp2_min = -18
tau_max = 0.3

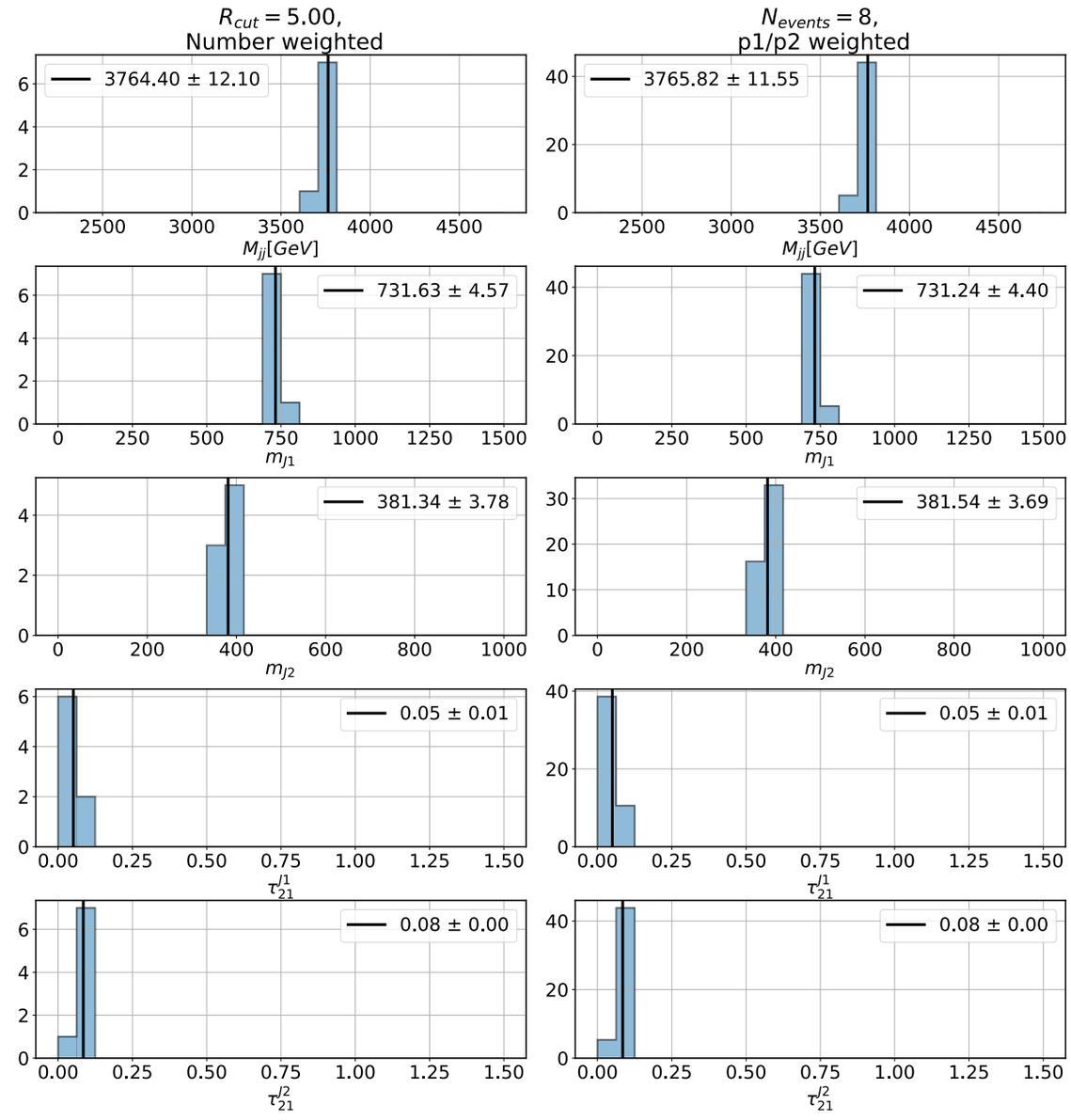
msig = 3766
delta_msig = 250

m1 = 730.
delta_m1 = 100
m2 = 370
delta_m2 = 100

BlackBox 1

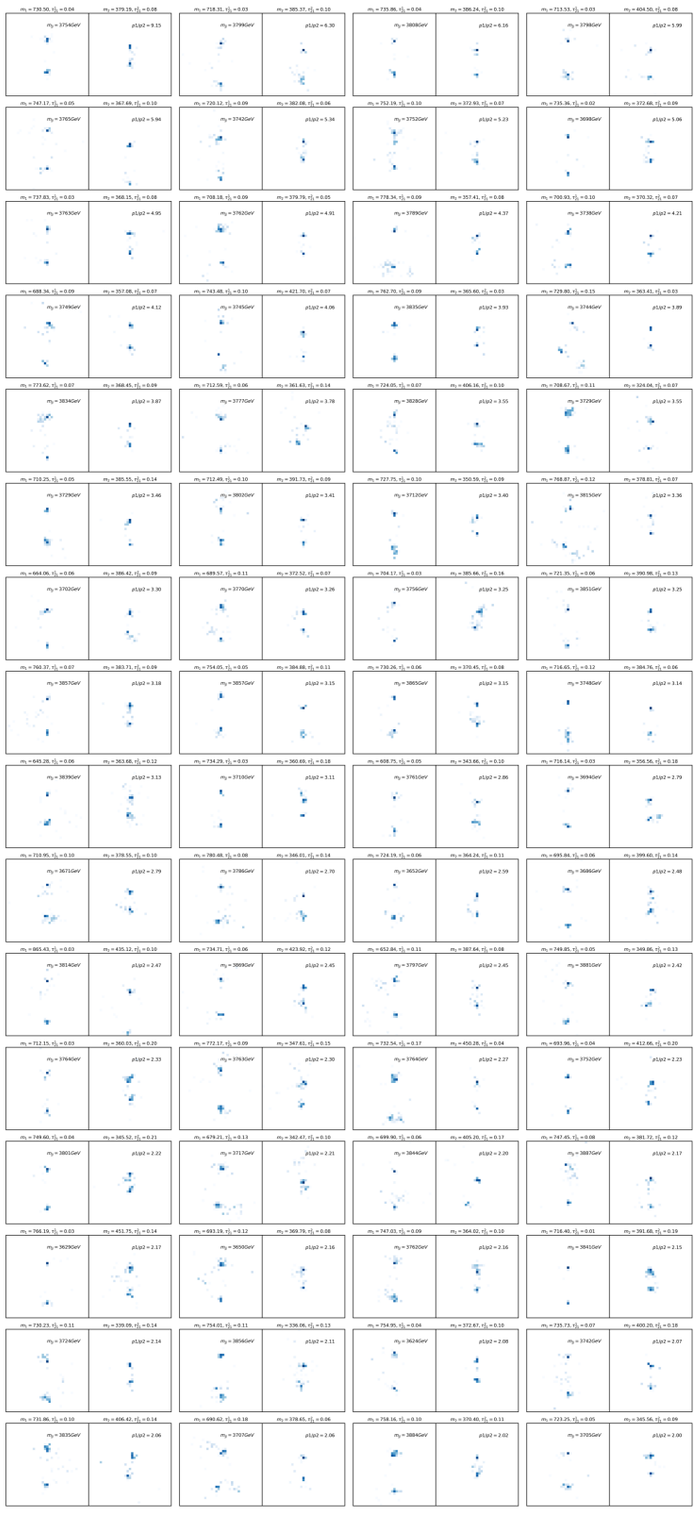
Reporting final parameters of events



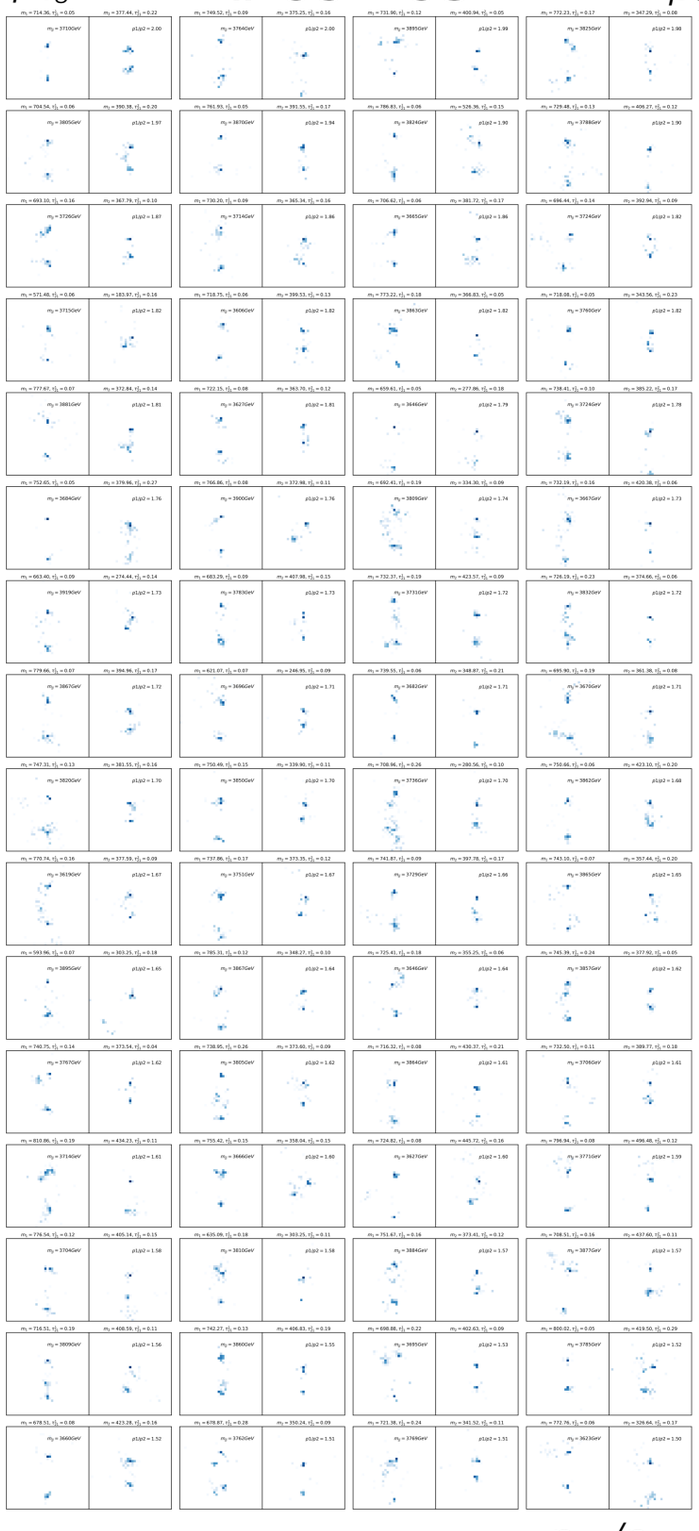


events

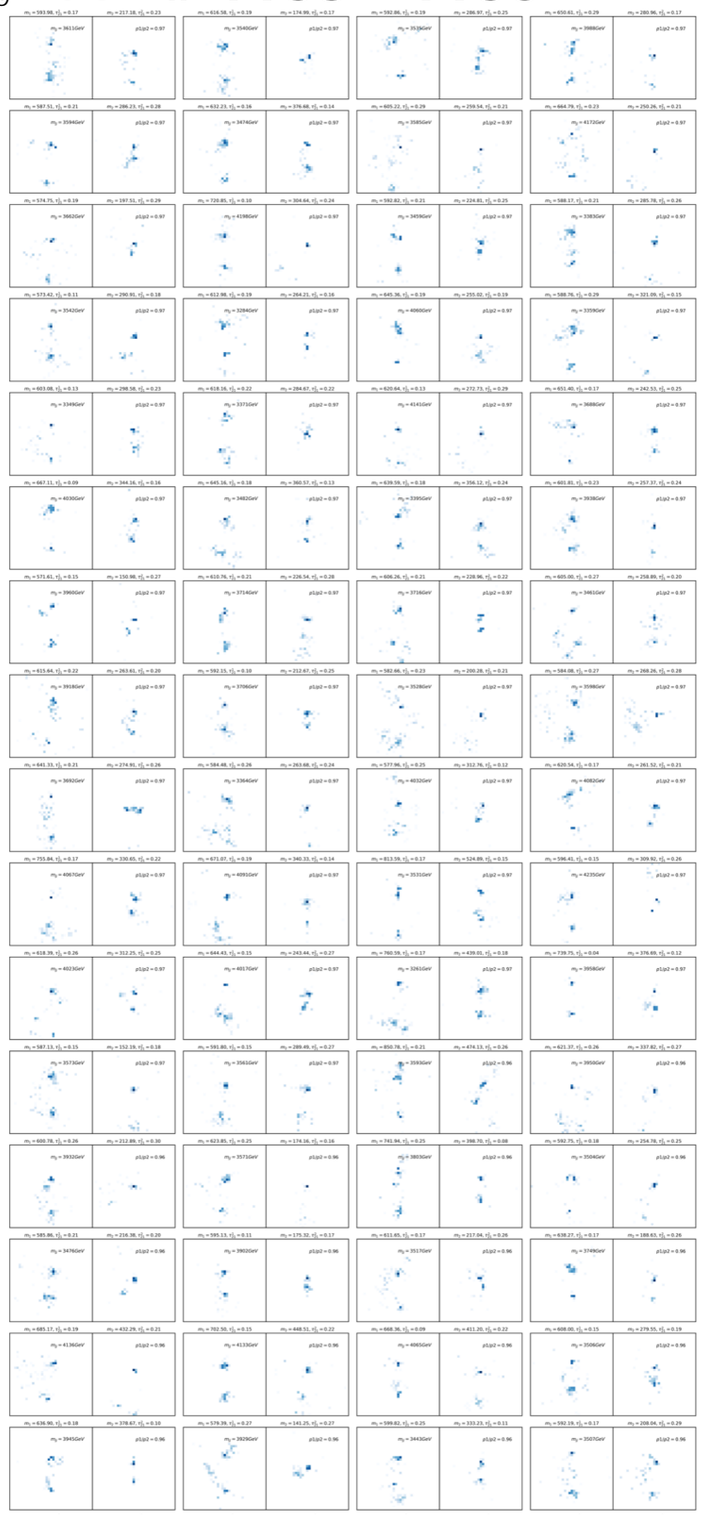
$p_{sig}/p_{bg} = 9.15$ **# 1 - 4**



$p_{sig}/p_{bg} = 2.0$ **# 65 - 68**



$p_{sig}/p_{bg} = 0.97$ **# 1153 - 1156**



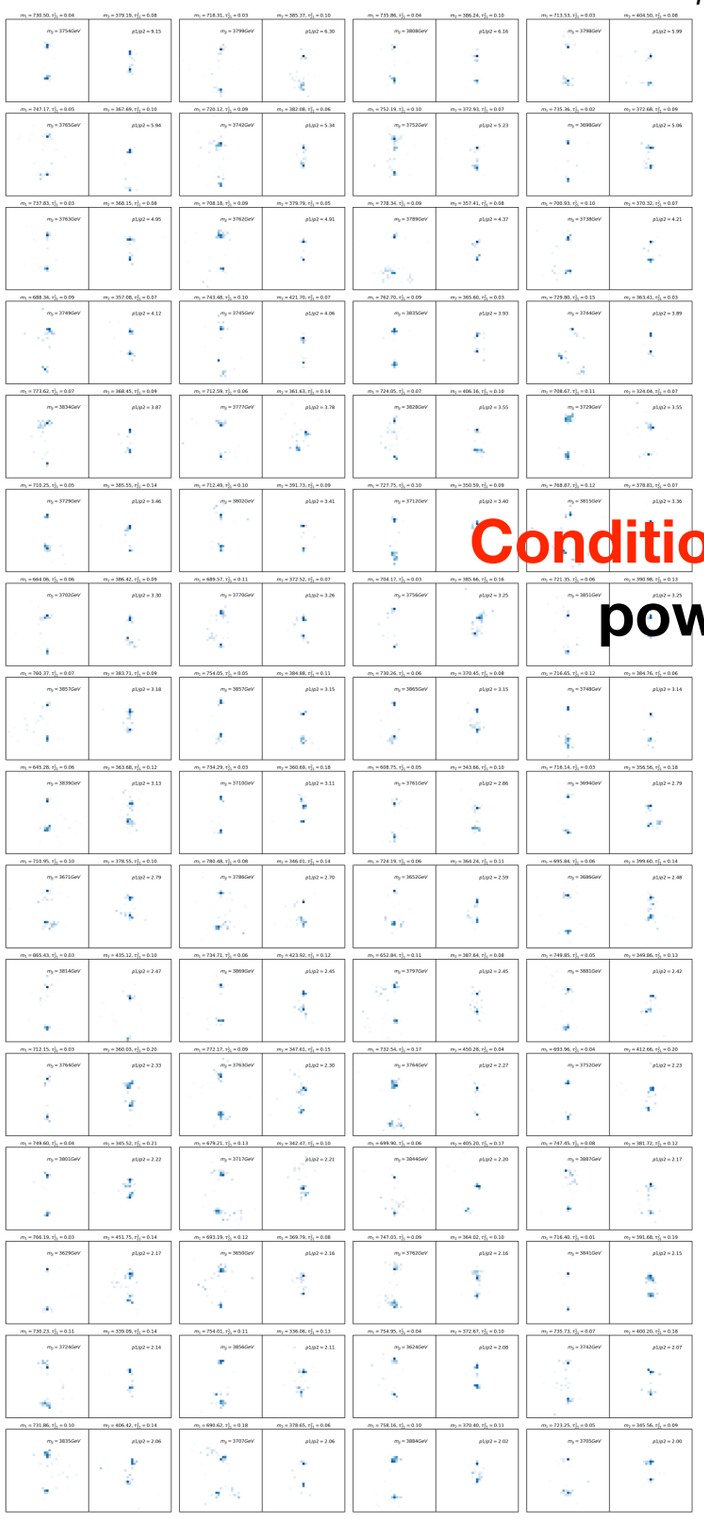
...

 # 61 - 64 $p_{sig}/p_{bg} = 2.0$ **# 125 - 128** $p_{sig}/p_{bg} = 1.5$ **# 1217 - 1220**

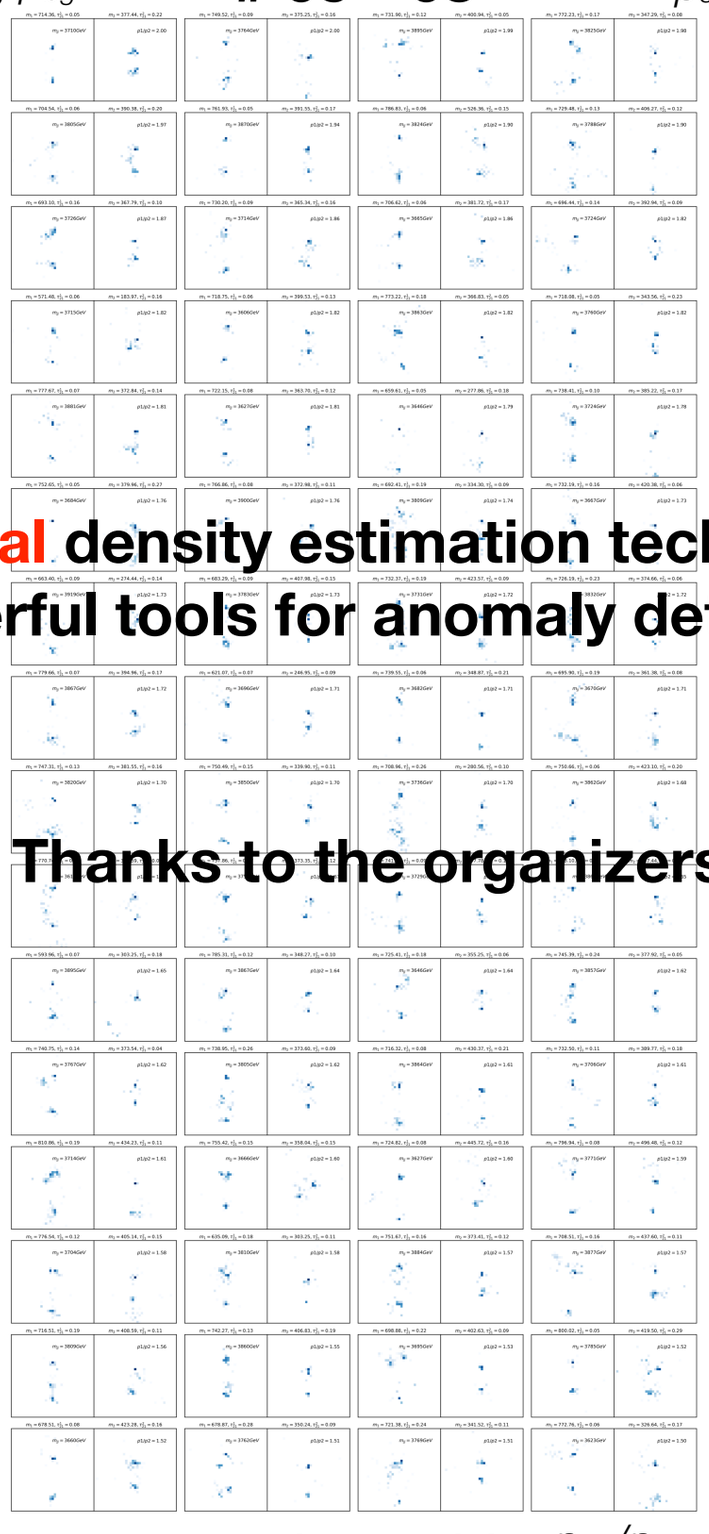
George Stein
BERKELEY CENTER *for* COSMOLOGICAL PHYSICS

events

$p_{sig}/p_{bg} = 9.15$ **# 1 - 4**



$p_{sig}/p_{bg} = 2.0$ **# 65 - 68**



$p_{sig}/p_{bg} = 0.97$ **# 1153 - 1156**

Conditional density estimation techniques are powerful tools for anomaly detection

...

Thanks to the organizers!

 # 61 - 64 $p_{sig}/p_{bg} = 2.0$

 # 125 - 128 $p_{sig}/p_{bg} = 1.5$

 # 1217 - 1220

George Stein
BERKELEY CENTER *for* COSMOLOGICAL PHYSICS